



# Storage over PCIe® Design and Validation Techniques

Isaac Livny

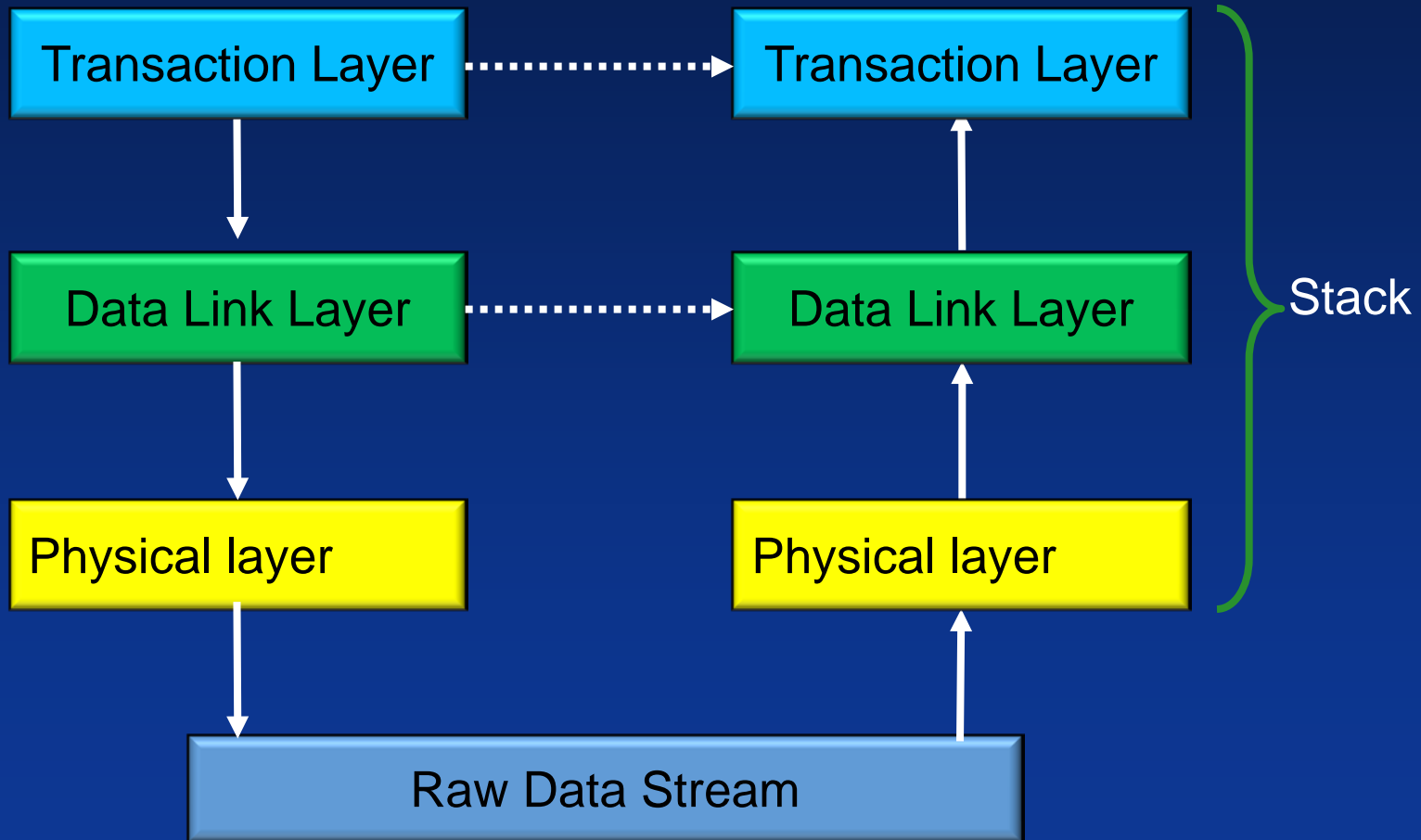
Field Applications Engineer  
Teledyne LeCroy Corporation



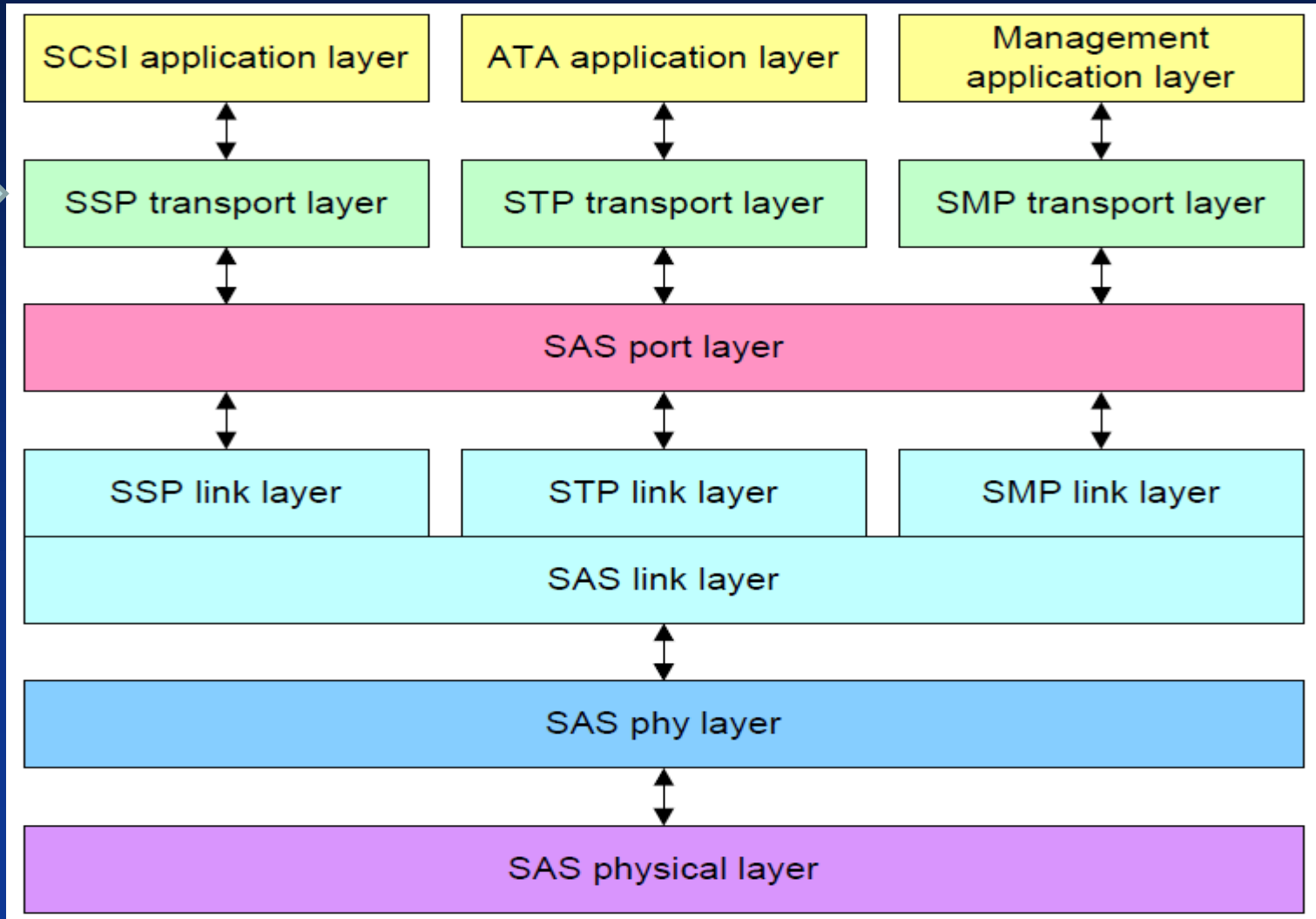
# Agenda

- Layered Protocol Stack convergence
- Storage over PCI Express<sup>®</sup> architectural description
- Command queue generation example
- Emulating an SSD controller
- Emulating an SSD host
- Command Validation

# PCIe® Layered Protocol Stack



# SAS Layered Protocol Stack



# Layered Protocols

## Support in Analysis Tools



- Hierarchical view display capability with multi layer expansion into sub-layers
- Multi-view capabilities
- Processing capability of upper layer through scripting to adopt to specification changes
- Tooltip feature to highlight specification details
- Performance and statistical analysis per instruction, by segment and overall trace
- Compacting of repetitive traffic
- Compacting of multiple 32 bit transactions into 64 bit upper layer commands

# PCIe Hierarchical View Split Transaction

Split Tra	0	R→	2.5	Cfg	CfgRd0	00:00100	RequesterID	000:03:0	CompleterID	003:00:0	Tag	0	TC	0	VC ID	0	DeviceID	003:00:0	Register	0x000	Status	SC
Device ID	0xF015	Vendor ID	0x10DF	Metrics	# LinkTras	2	Resp. time	556.000 ns	Latency	208.000 ns	Thrpt MB/s	6.861	Pld. Bytes	4	Time Stamp	0014 . 335 122 612 s						
Link Tra	16	R→	2.5	TLP	114	Cfg	CfgRd0	00:00100	Length	1	RequesterID	000:03:0	Tag	0	DeviceID	003:00:0	Register	0x000	1st BE	1111	VC ID	0
ExplicitACK	Packet #48668	Metrics	# Packets	2	Time Stamp	0014 . 335 122 612 s																
Packet	48665	R→	2.5	TLP	114	Cfg	CfgRd0	00:00100	Length	1	RequesterID	000:03:0	Tag	0	DeviceID	003:00:0	Register	0x000	1st BE	1111		
LCRC	0xA275BADB	Time Delta	208.000 ns	Time Stamp	0014 . 335 122 612 s																	
Packet	48668	R→	2.5	DLLP	ACK	AckNak_Seq_Num	114	CRC 16	0xF613	Time Delta	12.000 ns	Time Stamp	0014 . 335 122 820 s									
Link Tra	17	R→	2.5	TLP	4044	Cpl	CplID	10:01010	Length	1	RequesterID	000:03:0	Tag	0	CompleterID	003:00:0	Status	SC	BCM	0	Byte Cnt	4
LwrAddr	0x00	Device ID	0xF015	Vendor ID	0x10DF	VC ID	0	ExplicitACK	Packet #48670	Metrics	# Packets	2	Resp. time	336.000 ns	Pld. Bytes	4	Thrpt MB/s	11.353				
Time Stamp	0014 . 335 122 832 s																					
Packet	48669	R→	2.5	TLP	4044	Cpl	CplID	10:01010	Length	1	RequesterID	000:03:0	Tag	0	CompleterID	003:00:0	Status	SC	BCM	0		
Byte Cnt	4	LwrAddr	0x00	Device ID	0xF015	Vendor ID	0x10DF	LCRC	0xD977B772	Time Delta	332.000 ns	Time Stamp	0014 . 335 122 832 s									
Packet	48670	R→	2.5	DLLP	ACK	AckNak_Seq_Num	4044	CRC 16	0xC5AA	Time Delta	-504.000 ns	Time Stamp	0014 . 335 123 164 s									
Packet	48666	R→	2.5	SKIP	COM	SKIP Symbols	K28.5 K28.0 K28.0 K28.0	Idle	140.000 ns	Time Stamp	0014 . 335 122 660 s											
Packet	48667	R→	2.5	DLLP	UpdateFC-NP	VC ID	0	HdrFC	209	DataFC	134	CRC 16	0xFF36	Time Delta	556.000 ns	Time Stamp	0014 . 335 122 816 s					

# SAS Hierarchical View SCSI Command Decode

I/I	SCSI Cmd.	6 G	Source Address (H)	Destination Address (H)	Operation Code	Obsolete (H)	FUA_NV (H)	FUA (H)	DPO (H)	RDPROTECT (H)	Logical Block Address (H)	Group Number (H)	Transfer Length (H)
I1	1	6 G	500605B00273EFA0	5000C50009511682	0x28 : Read (10)	0	0	0	0	0	0000825B	00	0001
Control (H)	Payload Data , 512 Bytes			Task Attribute	Tag (H)	Status	LUN (H)	Metrics					
00	00 00 00 00 00 00 00 00 00 00 00 00 >>			0x0 : Simple	00D1	0x00 : Good	0000000000000000						

I/I	Transport	6 G	SSP Frame Type	Hashed Dest SAS Addr (H)	Hashed Src SAS Addr (H)	Changing Data Pointer (H)	ReTransmit (H)	Retry Data Frames (H)	TLR CONTROL (H)	Num of Fill Bytes (H)	Tag (H)
I1	1	6 G	0x06 : Command	13C1DF	FBB1DA	0	0	0	0	0	00D1
Target Port Transfer Tag (H)	Data Offset (H)	Info Unit (H)			CRC (H)	Handshake	Duration				
FFFF	00000000	00 >>			A641A6DA	0x0 : ACK	126 (ns)				

I/I	Link	6 G	Source SAS Address (H)	Destination SAS Address (H)	SSP Frame Type	Operation Code	Logical Block Address (H)	Transfer Length (H)	Task Attribute	Tag (H)	Link Data (H)
I1	6	6 G	500605B00273EFA0	5000C50009511682	0x06 : Command	0x28 : Read (10)	0000825B	0001	0x0 : Simple	00D1	
Relative Time	Duration										
6 (ns)	126 (ns)										

T1	Link	6 G	Target	RD	Relative Time	Duration
T1	7	6 G	ACK	----	233 (ns)	6 (ns)

T1	Transport	6 G	SSP Frame Type	Hashed Dest SAS Addr (H)	Hashed Src SAS Addr (H)	Changing Data Pointer (H)	ReTransmit (H)	Retry Data Frames (H)	TLR CONTROL (H)	Num of Fill Bytes (H)	Tag (H)
T1	2	6 G	0x01 : Data	FBB1DA	13C1DF	0	0	0	0	0	00D1
Target Port Transfer Tag (H)	Data Offset (H)	Data , 512 Bytes			CRC (H)	Handshake	Duration				
70E4	00000000	00 00 00 00 00 00 00 00 00 00 00 00 >>			A5B0CA75	0x0 : ACK	920 (ns)				

T1	Link	6 G	Source SAS Address (H)	Destination SAS Address (H)	SSP Frame Type	Tag (H)	Link Data (H)	Relative Time	Duration
T1	20	6 G	5000C50009511682	500605B00273EFA0	0x01 : Data	00D1		73 (ns)	920 (ns)

I/I	Link	6 G	Initiator	RD	Relative Time	Duration
I1	21	6 G	ACK	----	1.013 (us)	6 (ns)

T1	Transport	6 G	SSP Frame Type	Hashed Dest SAS Addr (H)	Hashed Src SAS Addr (H)	Changing Data Pointer (H)	ReTransmit (H)	Retry Data Frames (H)	TLR CONTROL (H)	Num of Fill Bytes (H)	Tag (H)
T1	3	6 G	0x07 : Response	FBB1DA	13C1DF	0	0	0	0	0	00D1
Target Port Transfer Tag (H)	Data Offset (H)	Info Unit (H)			CRC (H)	Handshake	Duration				
70E4	00000000	00000000000000000000000000000000 >>			33EBE8C1	0x0 : ACK	100 (ns)				



# USB Hierarchical View SCSI Command Decode

SCSI Op	ADDR	Tag	SCSI CDB	READ(10)	Logical Block Addr	Data	SBSU Passed	Time Stamp	Metrics	#Xfers
39	2	0x899A6DE8			0x00000028	4096 bytes	res=0x0	35 . 440 487 032		3
Transfer	H	Bulk	ADDR	ENDP	Mass Storage	CBSU In Len	SCSI CDB	READ(10)	Time Stamp	
139	S	OUT	2	2		0x00001000			35 . 440 487 032	
Transaction	H	OUT	ADDR	ENDP	T	Data	ACK	Time Stamp		
36515	S	0x87	2	2	1	31 bytes	0x4B	35 . 440 487 032		
ch0	Packet	H	H	OUT	ADDR	ENDP	CRC5	Pkt Len	Idle	Time Stamp
	160089	↓	S	0x87	2	2	0x01	8	200.660 ns	35 . 440 487 032
ch0	Packet	H	H	DATA1	Data	CRC16	Pkt Len	Idle	Time Stamp	
	160090	↓	S	0xD2	31 bytes	0x5D5B	40	299.330 ns	35 . 440 487 366	
ch0	Packet	↑	H	ACK	Pkt Len	Time	Time Stamp			
	160091	D	S	0x4B	8	244.818 μs	35 . 440 488 332			
Transfer	H	Bulk	ADDR	ENDP	Mass Storage	Bytes Transferred	Time	Time Stamp		
140	S	IN	2	1		4096	210.866 μs	35 . 440 733 150		
Transfer	H	Bulk	ADDR	ENDP	Mass Storage	SBSU Passed	Time Stamp			
141	S	IN	2	1		res=0x0	35 . 440 944 016			
Command Status Wrapper, GOOD status, Residue Length=0										
Transaction	H	IN	ADDR	ENDP	T	Data	ACK	Time Stamp		
36575	S	0x96	2	1	1	13 bytes	0x4B	35 . 440 944 016		
ch0	Packet	H	H	IN	ADDR	ENDP	CRC5	Pkt Len	Idle	Time Stamp
	160222	↓	S	0x96	2	1	0x18	10	333.330 ns	35 . 440 944 016
ch0	Packet	↑	H	DATA1	Data	CRC16	Pkt Len	Idle	Time Stamp	
	160223	D	S	0xD2	13 bytes	0xA8EA	22	367.330 ns	35 . 440 944 516	
ch0	Packet	H	H	ACK	Pkt Len	Time	Time Stamp			
	160224	↓	S	0x4B	6	181.882 μs	35 . 440 945 250			
SCSI Op	ADDR	Tag	SCSI CDB	READ(10)	Logical Block Addr	Data	SBSU Passed	Time	Time Stamp	
40	2	0x899A6DE8			0x00000030	4096 bytes	res=0x0	750.018 μs	35 . 441 127 132	



- High speed protocol convergence
- Storage over PCI Express architectural description
- Command queue generation example
- Emulating an SSD controller
- Emulating an SSD host
- Command Validation

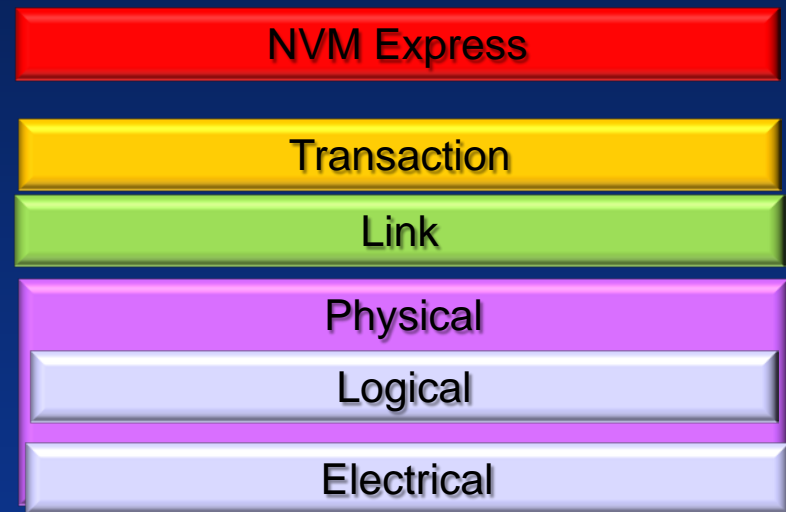


# NVM Express





- The NVMHCI Workgroup released the NVM Express 1.0 specification on March 1, 2011 and is available at [www.nvmexpress.org](http://www.nvmexpress.org)
- NVMe is a standardized high performance queuing interface and command set optimized for PCIe SSDs
- NVMe is scalable from client to enterprise applications





# NVM Express 1.0c Decodes



NVM		2.5	RequesterID	ASQS	ACQS	Time Delta	Time Stamp
0	R→	x4	000:00:0	383	383	512.000 ns	0046 . 434 680 292 s

**Admin Submission/completion Queue Size**

Submission Queue Size (SQS):  
Defines the size of the Admin Submission Queue in entries. The minimum size of the Admin Submission Queue is two entries. The maximum size of the Admin Submission Queue is 4096 entries. This is a 0's based value.

Completion Queue Size (CQS):  
Defines the size of the Admin Completion Queue in entries. The minimum size of the Admin Completion Queue is two entries. The maximum size of the Admin Completion Queue is 4096 entries. This is a 0's based value.

NVM		2.5	RequesterID	ASQB AddressHi	ASQB AddressLow	Time Delta	Time Stamp
1	R→	x4	000:00:0	0x00000001	0xEF224000	304.000 ns	0046 . 434 680 804 s

Admin Submission Queue Base (ASQB):  
Indicates the 64-bit physical address for the Admin Submission Queue. This address shall be memory page aligned (based on the value in CC.MPS). All Admin commands, including creation of additional Submission Queues and Com

**Admin Submission Queue Base Address**

**Tool Tips make it easy to understand each register item in a field**

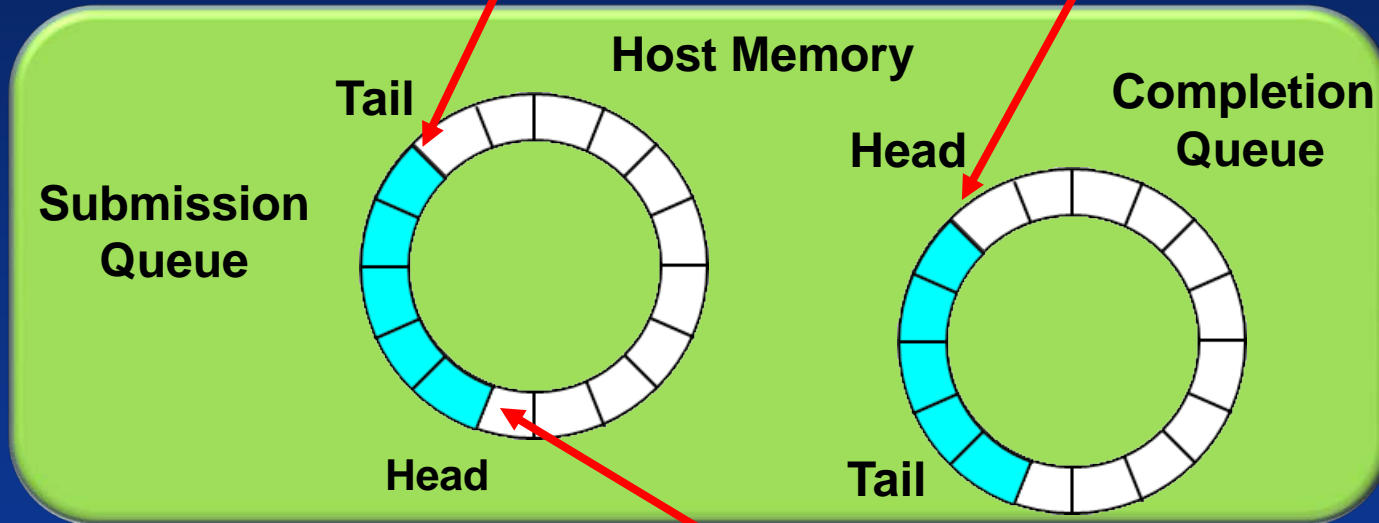
# Submission and Completion Queues

NVM	R→	2.5	RequesterID	Admin SQT QID = 0	Time Delta	Time Stamp
5		x4	000:00:0	0x0001	352.000 ns	0048 . 780 381 524 s

Submission Queue Tail

Completion Queue Head Pointer

NVM	R→	2.5	RequesterID	Admin CQH QID = 0	Time Delta	Time Stamp
22		x4	000:00:0	0x0003	271.860 μs	0078 . 827 032 124 s

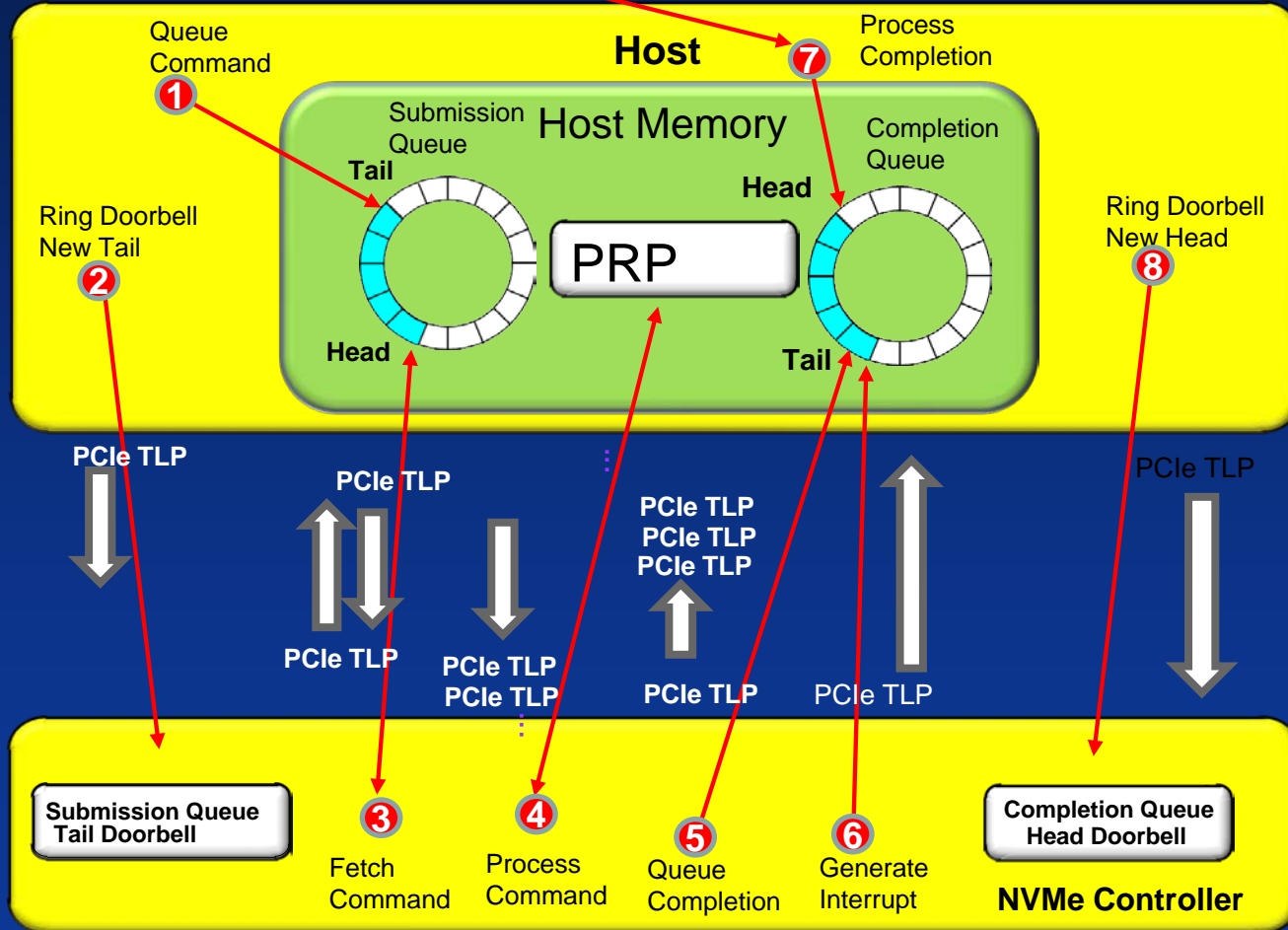


NVM	R←	2.5	RequesterID	Command Completion	SQHD	SQID	CID	P	ST	SC	SCT	M	DNR	Time Delta	Time Stamp
111		x4	007:00:0	0x00000000	0x0023	0x0000	0x0022	1		0x0000	0x00	0	0	26.752 μs	0030 . 998 377 752 s

Submission Queue Head Pointer

# Circular Queuing Interface

NVM	R	2.5	RequesterID	Command Completion	SQHD	SQID	CID	P	ST	SC	SCT	M	DNR	Time Delta	Time Stamp
111		x4	007:00:0	0x00000000	0x0023	0x0000	0x0022	1		0x0000	0x00	0	0	26.752 μs	0030 . 998 377 752 s





# Viewing the NVM Express 1.0c Initialization Phase

NVM	RequesterID	EN	CSS	MPS	AMS	SHN	IOSQES	IOCQES	Time Delta	Time Stamp																		
0	x4	000:00:0	1	NVM command set	0	b00	b00	0	152.000 ns	0078 . 759 536 468 s																		
1	x4	000:00:0	0	NVM command set	0	b00	b00	0	120.000 ns	0078 . 759 536 620 s																		
2	x4	000:00:0	008:00:0	CompleterID	MQES	CQR	AMS	TO	DSTRD	CSS	MPSMIN	MPSMAX	Time Delta	Time Stamp														
3	x4	000:00:0	127	127	168.000 ns	0078 . 759 536 732 s																						
4	x4	000:00:0	ASQB AddressHi	ASQB AddressLow	Time Delta	Time Stamp																						
5	x4	000:00:0	ACQB AddressHi	ACQB AddressLow	Time Delta	Time Stamp																						
6	x4	000:00:0	1	NVM command set	0	b00	b00	0	22.760 µs	0078 . 778 263 476 s																		
7	x4	000:00:0	008:00:0	CompleterID	RDY	CFS	SHST	Time Delta	Time Stamp																			
8	x4	000:00:0	0x0001	Admin SQT QID = 0	Time Delta	Time Stamp																						
9	x4	008:00:0	000:00:0	Admin Cmd	OPC	FUSE	CID	NSID	MPTR Hi	MPTR Low	PRP1 Hi	PRP1 Low	PRP2 Hi	PRP2 Low	CNS	Time Delta	Time Stamp											
10	x4	008:00:0	Identify CDS	VID	SSVID	SN	MN	FR	RAB	IEEE	MIC	MDTS	OACS	ACL	AERL	FRMW	LPA	ELPE	NPSS	AVSCC	SQES	CQES	NN	ONCS	FUSES	FNA		
11	x4	008:00:0	0x0000	0x111D	0x0000	0x0000	NVMIdentPid000000	0x000000005F45C6C3	0	0x0	0	0	0x0303	0	6	0	63	0	4	0	0	0	0	0x44660000	0x4	0x0	0	
12	0xFF	0x0	0	PSD0	MP	ENLAT	EXLAT	RRT	RRL	RWT	RWL	PSD1	MP	ENLAT	EXLAT	RRT	RRL	RWT	RWL	PSD2	MP	ENLAT	EXLAT	RRT	RRL	RWT	RWL	PSD3
13	0x0000	0x00000000	0x00000000	0x00	0x00	0x00	0x00	0x00	0x00	0x00	0x00	0x0000	0x00000000	0x00000000	0x00	0x00	0x00	0x00	0x00	0x0000	0x00000000	0x00000000	0x00	0x00	0x00	0x00	0x00	0x00
14	0x0000	0x00000000	0x00000000	0x00	0x00	0x00	0x00	0x00000000	0x00000000	0x00	0x00	VS	Data	Time Delta	Time Stamp													
15	0x0000	0x00000000	0x00000000	0x00	0x00	0x00	0x00000000	0x00000000	0x00	0x00	0x00	0x00	0x00	1023 bytes	1.628 ms	0078 . 794 167 248 s												
16	x4	008:00:0	0x00000000	Command Completion	SQHD	SQID	CID	P	ST	SC	SCT	M	DNR	Time Delta	Time Stamp													
17	x4	008:00:0	0x00000000	0x0001	0x0000	0x0000	1	ST	0x0000	0x00	0	0	0	26.720 µs	0078 . 795 794 960 s													

Physical Region Page pointer

Set submission and completion queues base addresses

Controller capabilities

Command Completion

Capabilities sent to Host memory

Courtesy SanDisk 2012

# Viewing the NVM Express 1.0c

## Identify Namespace Capabilities command

Namespace capabilities

NVM	RequesterID	CompleterID	Admin Cmd	OPC	FUSE	CID	NSID	MPTR Hi	MPTR Low	PRP1 Hi	PRP1 Low	PRP2 Hi	PRP2 Low	CNS	Time Delta					
14	R-	x4	008:00:0	000:00:0	Identify	b00	0x0001	0x00000001	0x00000000	0x00000000	0x00000001	0xEF269000	0x00000000	Namespace	55.752 μs					
Time Stamp 0078 . 809 588 752 s																				
NVM	RequesterID	Identify NS	NSZE	NCAP	NUSE	NSFEAT	NLBAF	FLBAS	MC	DPC	DPS	LBA0	MS	LBADS	RP	LBA1	MS	LBADS		
15	R-	x4	008:00:0	0x000000000007F6FD	0x000000000007F6FD	0x0000000000000000	0x03	0x00	0x05	0x00	Disabled	0x0000	0x09	Best	0x0000	0x0C				
RP	LBA2	MS	LBADS	RP	LBA3	MS	LBADS	RP	LBA4	MS	LBADS	RP	LBA5	MS	LBADS	RP	VS	Data	Time Delta	Time Stamp
Best	0x0010	0x09	Best	0x0010	0x0C	Best	0x0008	0x09	Best	0x0008	0x0C	Best	0x0008	0x0C	Best	3711 bytes	1.556 ms	0078 . 809 744 504 s		
NVM	RequesterID	Command Completion	QID	SQID	CID	P	ST	SC	SCT	M	DNR	Time Delta	Time Stamp							
16	R-	x4	008:00:0	0x00000000	0x0002	0x0000	0x0001	1	0x0000	0x00	0	0	26.736 μs	0078 . 811 300 960 s						

Command Completion





# Viewing the NVM Express 1.0c

## Creating an I/O queue and Read command example

Create IO Submission Queue

NVM	RequesterID	CompleterID	Admin Cmd	OPC	FUSE	CID	NSID	MPTR Hi	MPTR Low	PRP1 Hi	PRP1 Low	PRP2 Hi	PRP2 Low	QSIZE	QID	CQID	QPRIQ	PC
150	008:00:0	000:00:0	Create I/O SQ	b00	0x0024	0x00000000	0x00000000	0x00000000	0x00000000	0xEF31A000	0x00000000	0x00000000	0x00000000	1023	3	3	Medium	1
NVM	RequesterID	CompleterID	Command Completion	SQHD	SQID	CID	P	ST	SC	SCT	M	DNR	Time Delta	Time Stamp				
151	008:00:0	000:00:0	0x00000000	0x0025	0x0000	0x0024	1		0x0000	0x00	0	0	26.736 us	0048 . 897 517 192 s				

Courtesy SanDisk 2012

Data Transfer

Command Completion

Physical Region Page pointer

NVM	RequesterID	CompleterID	IO Cmd	OPC	FUSE	CID	NSID	MPTR Hi	MPTR Low	PRP1 Hi	PRP1 Low	PRP2 Hi	PRP2 Low	SLBA	LR	FUA	PRINFO	
195	008:00:0	000:00:0	Read	b00	0x0000	0x00000001	0x00000000	0x00000000	0x00000000	0x00000001	0xEED31000	0x00000000	0x00000000	0x00000000:00000000	0	0	0x0	
NLB	DSM	Incompressible	SR	AL	AF	EILBRT	ELBAT	ELBATM	Time Delta	Time Stamp								
0x0000		0	0	None	None	0x00000000	0x0000	0x0000	163.968 μs	0079 . 553 228 928 s								
NVM	RequesterID	CompleterID	CMD PRP	Addr Hi	Addr Lo	Data Len	Data	Time Delta	Time Stamp									
196	008:00:0	000:00:0	0x00000001	0xEED31000	0x00000200	127 quadlets		313.872 μs	0079 . 553 392 896 s									
NVM	RequesterID	CompleterID	Command Completion	SQHD	SQID	CID	P	ST	SC	SCT	M	DNR	Time Delta	Time Stamp				
197	008:00:0	000:00:0	0x00000000	0x0001	0x0010	0x0000	1		0x0000	0x00	0	0	26.720 μs	0079 . 553 706 768 s				

Courtesy SanDisk 2012

Data



# Viewing the NVM Express 1.0c NVMe Multiple Pointer Based Transactions

Physical Region Page (PRP)

NVM	Req	IO SQT QID = 1	RequesterID	CompleterID	IO Cmd	OPC	FUSE	CID	NSID	MPTR Hi	MPTR Low	PRP1 Hi	PRP1 Low	PRP2 Hi	PRP2 Low	SLBA	LR	FUA	PRINFO	NLB	DSM	Incompressib						
89	R	5.0	x8	000:00:0	0x000B																							
90	R	5.0	x8	129:00:0	000:31:1	Read	b00	0x000A	0x00000001	0x00000000	0x00000000	0x00000008	0x8FF41000	0x00000004	0x2FBFFB00	0x00000000:00000800	0	0	0x0	0x001F		0						
91	R	5.0	x8	129:00:0	000:31:1	PRP LIST PTR		0x00000008	0x2FBFFB00	0x00000020																		
												Data		Time Delta		Time Stamp												
												0: 8FF42000 00000008 8FF43000 00000008		168.000 ns		0009.438 507 150 s												
												4: 8FF44000 00000008 00000000 00000000																
92	R	5.0	x8	129:00:0		CMD PRP		Addr Hi	Addr Lo	Data Len	Data	Time Delta	Time Stamp															
								0x00000008	0x8FF41000	0x00001000	1023 quadlets	1.816 µs	0009.438 516 318 s															
93	R	5.0	x8	129:00:0		CMD PRP		Addr Hi	Addr Lo	Data Len	Data	Time Delta	Time Stamp															
								0x00000008	0x8FF42000	0x00001000	1023 quadlets	2.124 µs	0009.438 509 166 s															
94	R	5.0	x8	129:00:0		CMD PRP		Addr Hi	Addr Lo	Data Len	Data	Time Delta	Time Stamp															
								0x00000008	0x8FF43000	0x00001000	1023 quadlets	1.284 µs	0009.432 511 290 s															
95	R	5.0	x8	129:00:0		CMD PRP		Addr Hi	Addr Lo	Data Len	Data	Time Delta	Time Stamp															
								0x00000008	0x8FF44000	0x00001000																		
												Data																
												0: 00010000 00090008 00530052 00550054 00570056 00590058 005B005A 005D005C 00090008 00530052 00550054 00570056																
												12: 00190018 00210020 004B004A 004D004C 004F004E 00510050 00530052 00550054 00210020 004B004A 004D004C 004F004E																
												24: 00310030 00390038 00450044 00370036 00330032 003B003A 003D003C 003F003E 00390038 00450044 00370036 00330032																
												36: 00490048 00510050 005D005C 004F004E 004B004A 00530052 00550054 00570056 00510050 005D005C 004F004E 004B004A																
												48: 00410060 00490048 00550054 00470046 00430042 004B004A 004D004C 004F004E 00490048 00550054 00470046 00430042																
												60: 00590058 00810080 008D008C 007F007E 007B007A 00830082 00850084 00870086 00810080 008D008C 007F007E 007B007A																
												72: 00910090 00CD00CC 00CF00CE 00D100D0 00D300D2 00D500D4 00D700D6 00D900D8 00CD00CC 00CF00CE 00D100D0 00D300D2																
												84: 00A900A8 00C500C4 00C700C6 00C900C8 00CB00CA 00CD00CC 00B500B4 00B700B6 00C500C4 00C700C6 00C900C8 00CB00CA																
												96: 00C100C0 00C900C8 00D500D4 00C700C6 00C300C2 00C100C0 00C900C8 00D500D4 00C700C6 00C300C2 00C100C0 00C900C8																
												108: 00D900D8 00C100C0 00CD00CC 00DF00DE 005D005C 00110010 00130012 00150014 00C900C8 00CB00CA 00CD00CC 00CF00CE																
												120: 00D100D0 00D900D8 00D500D4 00E700D6 00550054 00290028 002B002C 002D002E ...																
96	R	5.0	x8	129:00:0		Command Completion	SQHD	SQID	CID	P	ST	SC	SCT	M	DNR	Time Delta	Time Stamp											
												0x00000000	0x000B	0x0001	0x000A	1		0x0000	0x00	0	0	48.000 ns	0009.438 514 894 s					

PRP List of pointers to memory addresses

Data



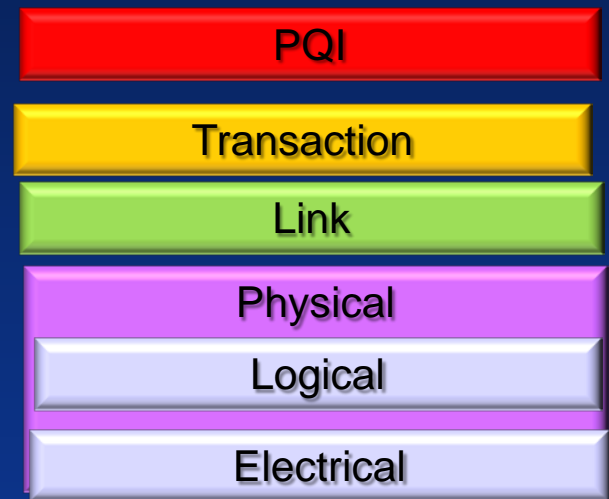
# PCIe Architecture Queuing Interface (PQI)



*SCSI Express*

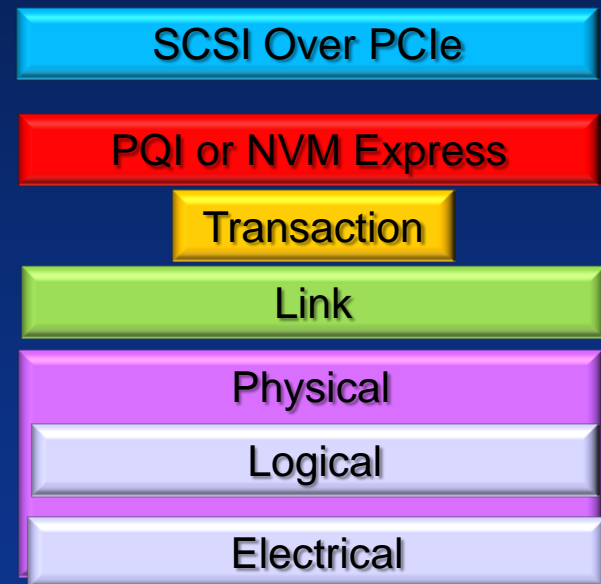
# PCIe Architecture Queuing Interface (PQI)

- T10 has developed this standard.
- Defines the transport methods for exchanging information between SCSI devices using a PCI Express interconnect
- Defines a queuing layer, potentially used by SOP
- Alternative to NVM Express
- Target and Initiator Support
- Targeting PCIe 3.0



# SCSI Over PCI Express(SOP/SOX)

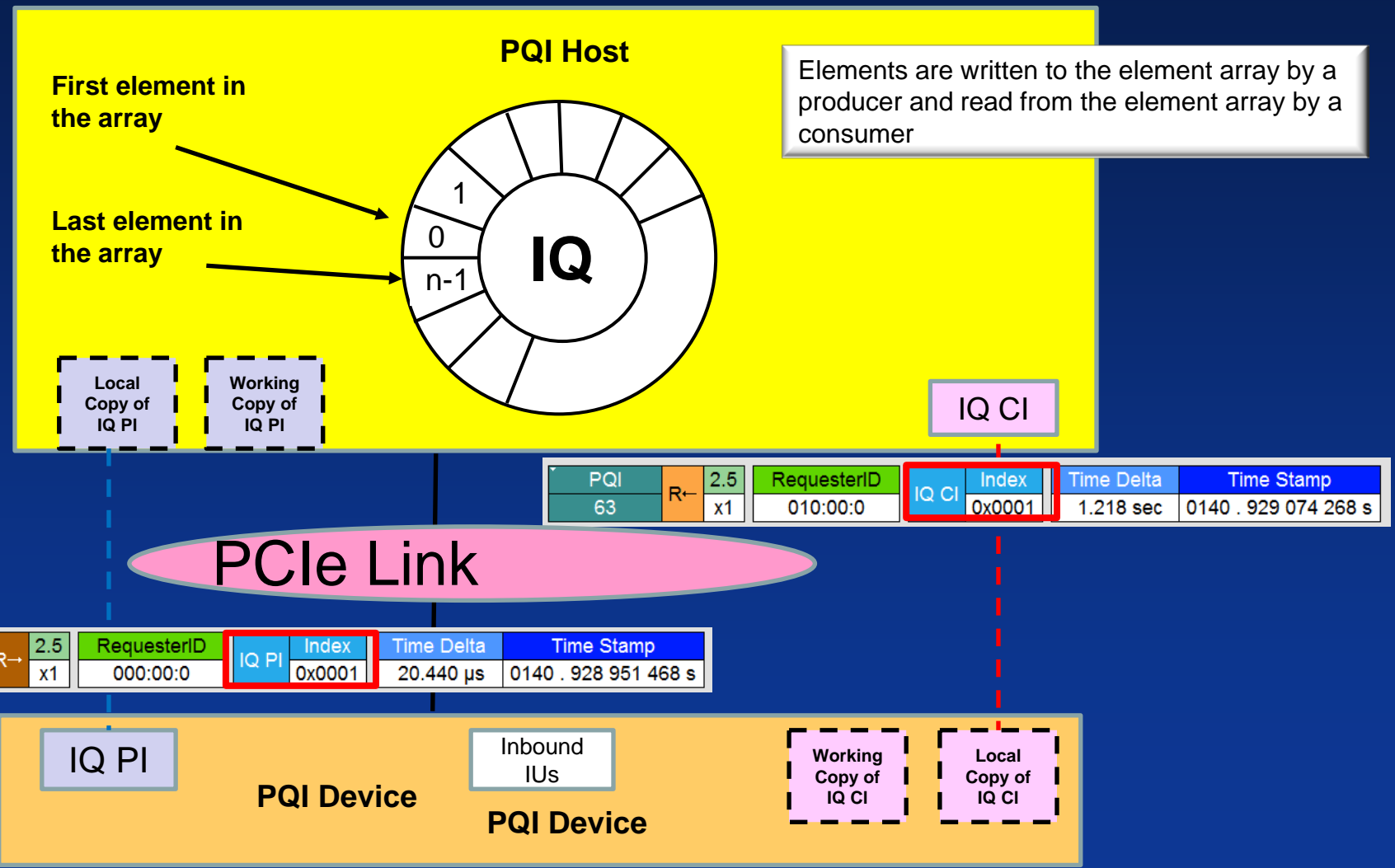
- Developed by T10 Committee
- Compliance with SCSI Architectural Model
- Proposed support for SOP target ports interfacing to flash devices, RAID controllers, and other SCSI peripheral device types
- Targeting PCIe 3.0



Preliminary  
Protocol Stack

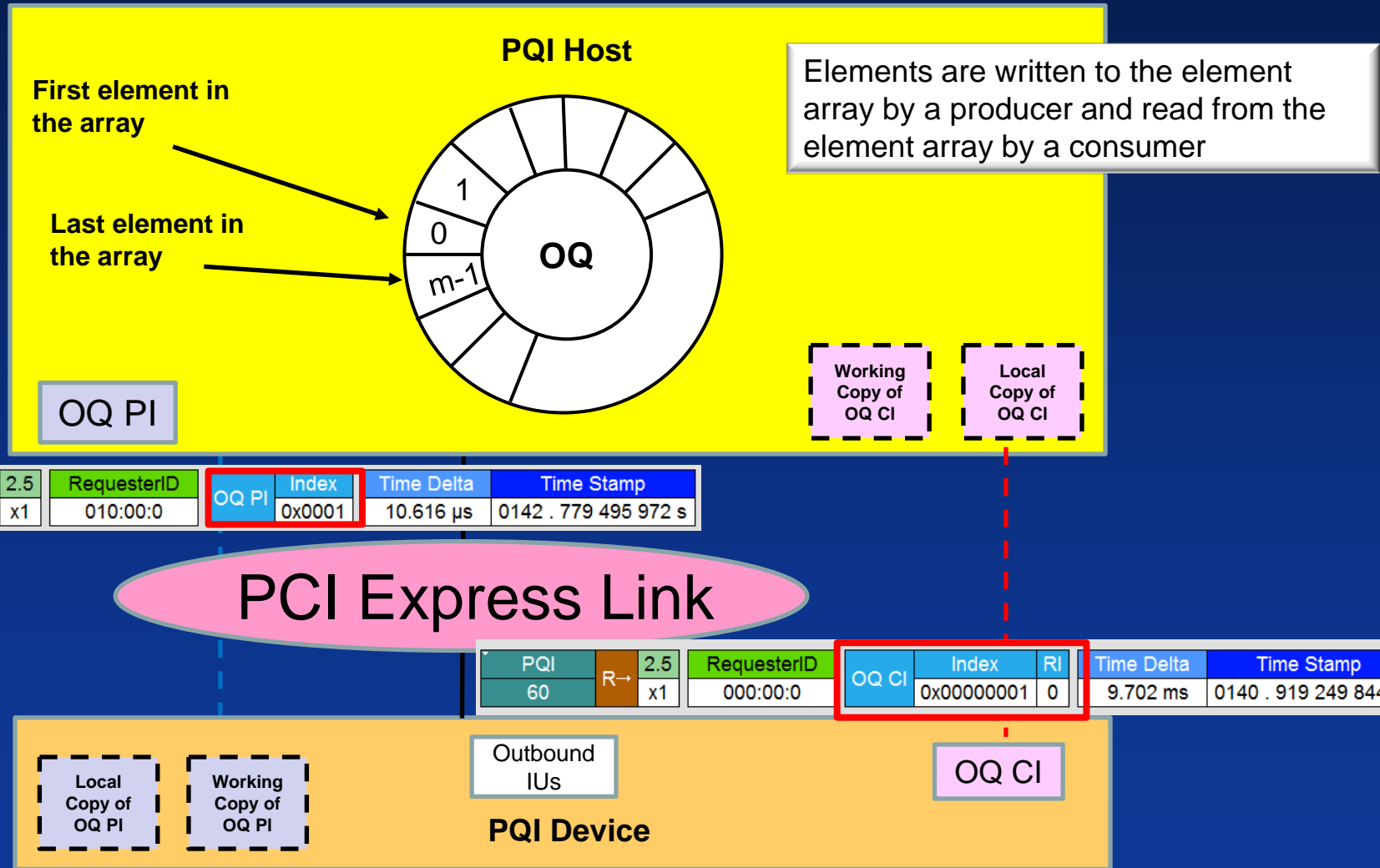


# SCSI Express SOP/PQI IQ(Inbound Queue)



# SCSI Express SOP/PQI

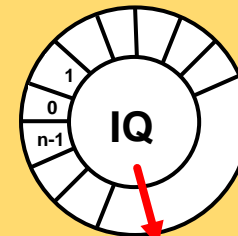
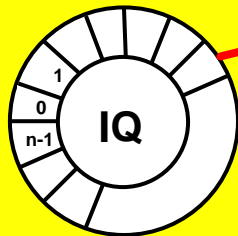
## OQ(Outbound Queue)



# Creating Administrative and Operational Queues

## Creating an Operational Inbound Queue

PQI	R+	2.5	RequesterID	CompleterID	Admin IU	IT	CF	IU LEN	ROQID	WA	Req ID	Op Code	IQ ID	IQEAA Hi	IQEAA Lo	IQCIA Hi
38		x1	010:00:0	000:00:0		0x60	0x00	0x003C	0x0000	0x0000	0x0004	Create Operational IQ	0x0000	0x00000000	0xBF70D000	0x00000000
IQCIA Lo	Num of Elements	Element Len	IMN	WFR	CC	MIN CT	MAX CT	Protocol	Time Delta	Time Stamp						
0xBF7089C8	0x00000040	0x0040	0x0000	0	0x0000	0x0000	0x0000	SOP	79.776 μs	0140 . 040 995 716 s						

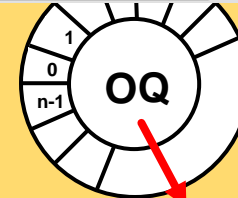
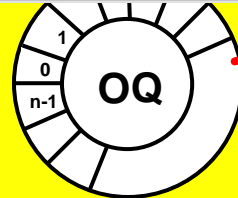


Response

PQI	R+	2.5	RequesterID	Admin OU	IT	CF	IU LEN	Reserved	WA	Req ID	OP Code	Status	QPI Offset Hi	IQPI offset Lo	Time Delta
39		x1	010:00:0		0xE0	0x00	0x003C	0x0000	0x0000	0x0004	Create Operational IQ	Success	0x00000000	0x00001010	22.792 μs

## Creating an Operational Outbound Queue

PQI	R+	2.5	RequesterID	CompleterID	Admin IU	IT	CF	IU LEN	ROQID	WA	Req ID	Op Code	OQ ID	OQEAA Hi	OQEAA Lo	OQPIA Hi
32		x1	010:00:0	000:00:0		0x60	0x00	0x003C	0x0000	0x0000	0x0000	Create Operational OQ	0x0000	0x00000000	0xBF712000	0x00000000
OQPIA Lo	Num of Elements	Element Len	IMN	WFR	CC	MIN CT	MAX CT	Protocol	Time Delta	Time Stamp						
0xBF708BD8	0x00000040	0x0040	0x0003	0	0x0000	0x0000	0x0000	SOP	101.840 μs	0140 . 038 995 284 s						



Response

PQI	R+	2.5	RequesterID	Admin OU	IT	CF	IU LEN	Reserved	WA	Req ID	OP Code	Status	OQCI Offset Hi	OQCI offset Lo	Time Delta
33		x1	010:00:0		0xE0	0x00	0x003C	0x0000	0x0000	0x0003	Create Operational OQ	Success	0x00000000	0x00001038	22.816 μs

Admin Queue

Operational Queue





# SCSI Express Initialization

PQI	RequesterID	CompleterID	PAF	STATUS	Time Delta	Time Stamp																				
5	x1	000:01:0	010:00:0	Idle	1.551 ms	0140 . 032 247 892 s																				
6	x1	000:01:0	010:00:0	ADMIN Queue Parameter	ADMIN IQ ELEMENTS	ADMIN OQ ELEMENTS	INTERRUPT MESSAGE	Time Delta	Time Stamp																	
7	x1	000:01:0	010:00:0	CAPABILITY	MAX ADMIN IQ	MAX ADMIN OQ	AIQE LEN	AOQE LEN	Time Delta	Time Stamp																
8	x1	000:01:0	010:00:0	CAPABILITY	MAX ADMIN IQ	MAX ADMIN OQ	AIQE LEN	AOQE LEN	Time Delta	Time Stamp																
9	x1	000:00:0	010:00:0	ADMIN Queue Parameter	ADMIN IQ ELEMENTS	ADMIN OQ ELEMENTS	INTERRUPT MESSAGE	Time Delta	Time Stamp																	
10	x1	000:00:0	010:00:0	ADMIN IQ	Offset Hi	Offset Lo	Time Delta	Time Stamp																		
11	x1	000:00:0	010:00:0	ADMIN IQ CI	Offset Hi	Offset Lo	Time Delta	Time Stamp																		
12	x1	000:00:0	010:00:0	ADMIN OQ	Offset Hi	Offset Lo	Time Delta	Time Stamp																		
13	x1	000:00:0	010:00:0	ADMIN OQ PI	Offset Hi	Offset Lo	Time Delta	Time Stamp																		
14	x1	000:01:0	010:00:0	PAF	STATUS	Time Delta	Time Stamp																			
15	x1	000:00:0	010:00:0	PAF	STATUS	Time Delta	Time Stamp																			
16	x1	000:01:0	010:00:0	PAF	STATUS	Time Delta	Time Stamp																			
17	x1	000:01:0	010:00:0	ADMIN IQ PI	Offset Hi	Offset Lo	Time Delta	Time Stamp																		
18	x1	000:01:0	010:00:0	ADMIN OQ CI	Offset Hi	Offset Lo	Time Delta	Time Stamp																		
19	x1	000:00:0	010:00:0	RequesterID	Index	Time Delta	Time Stamp																			
20	x1	010:00:0	000:00:0	Admin IU	IT	CF	IU LEN	ROQID	WA	Req	Op Code	IQ ID	OQEAA Hi	OQEAA Lo	OQPIA Hi	OQPIA Lo	Num of Elements	Element Len	IMN	WFR	CC	MIN CT	MAX CT	Protocol	Time Delta	
											Create Operational OQ	0x0000	0x00000000	0xBF70E000	0x00000000	0xBF708BC3	0x00000040	0x0040	0x0001	0	0x0000	0x0000	0x0000	SOP	100.680 μs	
21	x1	010:00:0	000:00:0	Admin OQ	IT	CF	IU LEN	Reserved	WA	Req ID	OP Code	Status	OQCI Offset Hi	OQCI offset Lo	Time Delta	Time Stamp										
											Create Operational OQ	Success	0x00000000	0x00001018	22.792 μs	0140 . 035 085 540 s										

Creating Administrative IQ and OQ Queues

Creating Operational OQ Queue



# SCSI Express SOP Transfer Packet

Advancing Producer in Inbound Queue

PQI	R	2.5	RequesterID	IQ PI	Index	Time Delta	Time Stamp
55	R	x1	000:00:0	0x0001	21.304 µs	0140 . 113 039 492 s	

PQI	R	2.5	RequesterID	CompleterID	SOP IU	IT	CF	IU LEN	ROQID	WA	Req ID	Data Dir	PARTIAL	DATA LEN	SCSI CDB	OP Code	SA	CDB Info	LBA	CDB Info	LEN	CDB Info	CONTROL
56	R	x1	010:00:0	000:00:0	0x10	0x00	0x003C	0x0003	0x0000	0x0000	0x0007	3	0	0x00000010	0xA0	0x0	0x0	0x00000000	0x00001000	0x00000000	0x00	0x00000000	

Advancing Consumer in Inbound Queue

PQI	R	2.5	RequesterID	IQ CI	Index	Time Delta	Time Stamp
57	R	x1	010:00:0	0x0001	190.879 ms	0140 . 113 162 956 s	

Link Tra	R	2.5	TLP	Msg	MsgD	Msg Routing	Length	RequesterID	Tag	Message Code	VID	Data	VC ID	Explicit ACK	Metrics	# Packets	Time Delta	Time Stamp
9610	R	x1	851	011:10011	Broadcast	1	000:00:0	0	Vendor_Defined_Type1	0x8086	1 dword	0	Packet #18600		2	135.073 ms	0140 . 604 041 748 s	

Link Tra	R	2.5	TLP	Mem	MWrr(32)	Length	RequesterID	Tag	Address	1st BE	Last BE	Data	VC ID	Explicit ACK	Metrics	# Packets	Time Delta	Time Stamp
9611	R	x1	568	010:00000	4	010:00:0	0	14281120	1111	1111	4 dwords	0	Packet #18602		2	180.095 ms	0140 . 739 115 076 s	

Advancing Producer in Outbound Queue

PQI	R	2.5	RequesterID	OQ PI	Index	Time Delta	Time Stamp			
58	R	x1	010:00:0	0x90	0x00	0x000C	0x0003	0x0000	0x0007	0x0007
59	R	x1	010:00:0	0x0001	10.664 µs	0140 . 919 233 540 s				

Link Tra	R	2.5	TLP	Mem	MWrr(32)	Length	RequesterID	Tag	Address	1st BE	Last BE	Data	VC ID	Explicit ACK	Metrics	# Packets	Time Delta	Time Stamp
9614	R	x1	571	010:00000	1	010:00:0	0	FEE22000	1111	0000	1 dword	0	Packet #18608		2	5.640 µs	0140 . 919 244 204 s	

PQI	R	2.5	RequesterID	OQ CI	Index	Time Delta	Time Stamp
60	R	x1	000:00:0	0x00000001	9.702 ms	0140 . 919 249 844 s	

Advancing Consumer in Outbound Queue

Courtesy SanDisk 2012



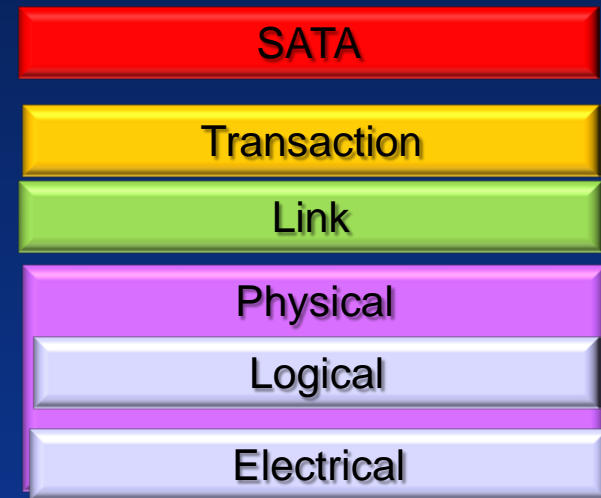
# SATA Express



# *SATA Express*

# SATA Express

- The Serial ATA International Organization (SATA-IO) developed the specification
- This protocol combines the SATA AHCI software specification with the PCIe host interface
- SATA Express enables new devices to be developed that utilize the faster PCIe interface and maintain compatibility with a broad base of existing SATA applications
- Data Rate Support
  - PCIe 2.x at x2 link for 8GT/s data rate
  - PCIe 3.0 at x2 link for a 16GT/s data rate



# AHCI HBA Registers

File Setup Record Generate Report Search View Tools Window Help

AHCI	RequesterID	CompleterID	CAP	S64A	SNCQ	SSNTF	SMPS	SSS	SALP	SAL	SCLO	ISS	SAM	SPM	FBSS	PMD	SSC	PSC	NCS	CCCS	EMS	SXS	NP	Time Delta	Time Stamp
0	5.0 x1	000:00:0 004:00:0	1	1	0	0	0	0	0	0	0	6.0 Gbps	0	0	0	1	0	0	31	0	0	0	7	22.976 µs	0008 . 574 012 896 s
1	5.0 x1	000:00:0 004:00:0	GHC	AE	Reserved	MRS	MRSM	IE	HR	Time Delta		Time Stamp													
2	5.0 x1	000:00:0	GHC	AE	Reserved	MRS	MRSM	IE	HR	Time Delta		Time Stamp													
3	5.0 x1	000:00:0	GHC	AE	Reserved	MRS	MRSM	IE	HR	Time Delta		Time Stamp													
4	5.0 x1	000:00:0	GHC	AE	Reserved	MRS	MRSM	IE	HR	Time Delta		Time Stamp													
5	5.0 x1	000:00:0	GHC	AE	Reserved	MRS	MRSM	IE	HR	Time Delta		Time Stamp													
6	5.0 x1	000:00:0	GHC	AE	Reserved	MRS	MRSM	IE	HR	Time Delta		Time Stamp													
7	5.0 x1	000:00:0	GHC	AE	Reserved	MRS	MRSM	IE	HR	Time Delta		Time Stamp													
8	5.0 x1	000:00:0	GHC	AE	Reserved	MRS	MRSM	IE	HR	Time Delta		Time Stamp													
9	5.0 x1	000:00:0	GHC	AE	Reserved	MRS	MRSM	IE	HR	Time Delta		Time Stamp													
10	5.0 x1	000:00:0	GHC	AE	Reserved	MRS	MRSM	IE	HR	Time Delta		Time Stamp													
11	5.0 x1	000:00:0	GHC	AE	Reserved	MRS	MRSM	IE	HR	Time Delta		Time Stamp													
12	5.0 x1	000:00:0	GHC	AE	Reserved	MRS	MRSM	IE	HR	Time Delta		Time Stamp													
13	5.0 x1	000:00:0	GHC	AE	Reserved	MRS	MRSM	IE	HR	Time Delta		Time Stamp													
14	5.0 x1	000:00:0	GHC	AE	Reserved	MRS	MRSM																		
15	5.0 x1	000:00:0	IS	Interrupt Pending Status																					
16	5.0 x1	000:00:0	IS	Interrupt Pending Status																					

CAP: Host Capabilities  
 GHC: Global Host Control  
 IS: Interrupt Status  
 PI: Ports Implemented

**HBA Memory Registers**

1. Port Control
2. Generic Host Control(GHC)

**AHCI PCIe Configuration Space Registers**

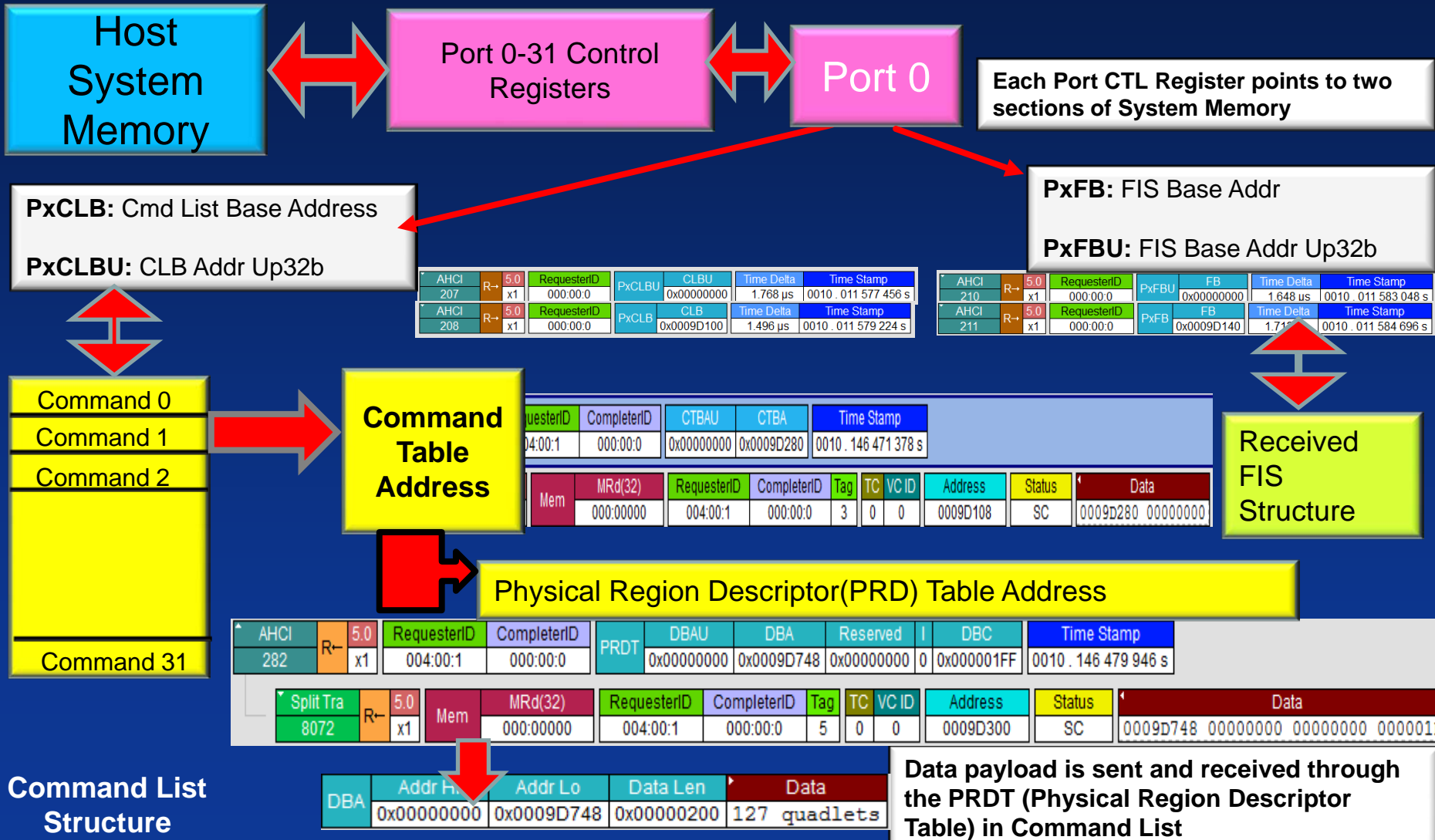
# SATA Express Port Control Setup

AHCI	R→	5.0	RequesterID	PxCMD	ICC	ASP	ALPE	DLAE	ATAPI	APSTE	FBSCP	ESP	CPD	MPSP	HPCP	PMA	CPS	CR	FR	MPSS	CCS	FRE	CLO	POD	SUD	ST	Time Delta	Time Stamp			
203	R→	5.0	000:00:0	PxCMD	No-Op / Idle	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	1.520 μs	0010 . 002 648 088 s			
204	R→	5.0	000:00:0	CompleterID	PxCMD	No-Op / Idle	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	0	812.776 μs	0010 . 002 649 608 s
205	R→	5.0	000:00:0	PxCMD	No-Op / Idle	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1.688 μs	0010 . 003 462 384 s		
206	R→	5.0	000:00:0	CompleterID	PxCMD	No-Op / Idle	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	8.113 ms	0010 . 003 464 072 s
207	R→	5.0	000:00:0	PxCLBU	CLBU	0x00000000																									
208	R→	5.0	000:00:0	PxCLB	CLB	0x0009D100																									
209	R→	5.0	000:00:0	CompleterID	PxCLB	CLB	0x0009D100																								
210	R→	5.0	000:00:0	PxFBU	FB	0x00000000																									
211	R→	5.0	000:00:0	PxFB	FB	0x0009D140																									
212	R→	5.0	000:00:0	CompleterID	PxFB	FB	0x0009D140																								

Port control address of Command list setup

Port control address of Received FIS setup

# AHCI System Memory



- Layered Protocol Stack convergence
- Storage over PCI Express architectural description
- Command queue generation example
- Emulating an SSD controller
- Emulating an SSD host
- Command Validation





# NVMe Device script Submission Queue Setup

NVM 0	R→	2.5 x4	RequesterID 000:00:0	AQA	ASQS 383	ACQS 383	Time Stamp 0029.315690784 s							
Link Tra 11783	R→	2.5 x4	TLP 2237	Mem	MWr(32) 010:00000	Length 1	RequesterID 000:00:0	Tag 10	Address B1A00024	1st BE 1111	Last BE 0000	Data 017F017F	VC ID 0	Exp Pack
NVM 1	R→	2.5 x4	RequesterID 000:00:0	ASQ	ASQB AddressHi 0x00000000	ASQB AddressLow 0x7F55A000	Time Stamp 0029.315690956 s							
Link Tra 11784	R→	2.5 x4	TLP 2238	Mem	MWr(32) 010:00000	Length 2	RequesterID 000:00:0	Tag 11	Address B1A00028	1st BE 1111	Last BE 1111	Data 7F55A000 00000000		
NVM 2	R→	2.5 x4	RequesterID 000:00:0	ACQ	ACQB AddressHi 0x00000000	ACQB AddressLow 0x7F561000	Time Stamp 0029.315691120 s							
Link Tra 11785	R→	2.5 x4	TLP 2239	Mem	MWr(32) 010:00000	Length 2	RequesterID 000:00:0	Tag 8	Address B1A00030	1st BE 1111	Last BE 1111	Data 7F561000 00000000		

- The host creates a command for execution within the appropriate Submission Queue
- Admin submission queue base register 0x7F55A000 is written to controller register, later used by controller to fetch the read command from host memory

```
# NVM 0
Wait=TLP {
    TLPTType=MWr32
    Length = 1 }
#NVM 1
Wait=TLP {
    TLPTType=MWr32
    Length = 2 }
```



# NVMe Device script Command fetch

NVM	2.5	RequesterID	CompleterID	Admin Cmd	OPC	FUSE	CID	NSID	MPTR Hi	MPTR Low	PRP1 Hi	PRP1 Low	PRP2 Hi	PRP2 Low	CNS	Time Stamp
8	R-	x4	007:00:0	000:00:0	Identify	b00	0x0000	0x00000000	0x00000000	0x00000000	0x00000000	0x7F5FA000	0x00000000	0x00000000	Controller	0030.972.808.984 s

Split Tra	2.5	Mem	MRd(32)	RequesterID	CompleterID	Tag	TC	VC ID	Address	Status	Data
5845	R-	x4	000:00000	007:00:0	000:00:0	0	0	0	7F55A000	SC	0: 00000006 00000000 00000000 00000000 00000000 00000000 7F5FA000 00000000 8: 00000000 00000000 00000001 00000000 00000000 00000000 00000000 00000000

NVM	2.5	RequesterID	Identify CDS	VID	SSVID	SN	MN	FR	RAB	IEEE	MIC	MDTS	OACS	ACL	AERL	FRMW	LPA	ELPE	NPSS	AVSOC	SQES	CQES	NN	ONCS	FUSES	FNA	VWC	AWUN	AWUPF	N
9	R-	x4	007:00:0	0x111D	0x0000				0	0x0	0	0	0x0000	255	0	4	0	0	0	0	102	68	0x00000004	0x0000	0x0000	0	0	0x00FF	0x00FF	N

EXLAT	RRT	RRL	RWT	RWL	VS	Data	Time Stamp
0x00000000	0x00	0x00	0x00	0x00		1023 bytes	0030.972.981.232 s

Link Tra	2.5	TLP	Mem	MW(32)	Length	RequesterID	Tag	Address	1st BE	Last BE	Data	VC ID	ExplicitACK	Metrics	# Packets	Time Delta	Time Stamp
11828	R-	x4	2128	010:00000	64	007:00:0	0	7F5FA000	1111	1111	64 dwords	0	Packet#23593		2	79.632 us	0030.972.981.232 s

- The controller fetches the command(s) in the Submission Queue from host memory
- The Admin submission queue base address register 0x7F55A000 is read from implemented memory resource through field substitution

```

packet = TLP {
  TLPType = MRd64
  Length = 0x10
  RequesterId = (8:0:0)
  Tag = 0x0
  LastDwBe = 0xF
  FirstDwBe = 0xF
  AddressHi = ( FROM_MEM32_A, 0x28 )
              # 0x1
  AddressLo = ( FROM_MEM32_A, 0x2C )
              # 0xEF224000 }
  
```

# Agenda

- High speed protocol convergence
- Storage over PCI Express architectural description
- Command queue generation example
- Emulating an SSD controller
- Emulating an SSD host
- Command Validation

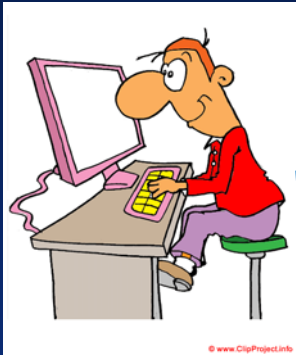
# Testing NVM Express

- NVMe Registers emulation
  - Setup Admin Queue
  - Doorbell Registers
- Admin Commands
  - Delete I/O Submission Queue
  - Create I/O Submission Queue
  - Delete I/O Completion Queue
  - Create I/O Completion Queue
  - Get Log Page
  - Identify
  - Abort
  - Set Features
  - Get Features
  - Format NVM
  - Extensible for Vendor Specific Commands



Summit Z3-16 Protocol Exerciser

# Pre-Silicon Device Emulation



RTL Design Phase  
Denali, Synopsys

RTL Test Vectors

CATC Trace

Z3 Exerciser  
Generation Script

SimPass

PE Tracer  
Export to  
Generation  
file



# Loading Config Space and Implementing Memory Space

**Write Address Space**

	File Path:	Offset (bytes):	Size (bytes):
<input checked="" type="checkbox"/> Cfg :	cripts\sys_device_cfg_space_w_msi_x_1_bar		
<input type="checkbox"/> Mem64 :		0x00000000	0x20000000
<input type="checkbox"/> Mem32 A :		0x00000000	0x08000000
<input checked="" type="checkbox"/> Mem32 B :	F:\Data_All\NVMe_Emulation\identify1	0x00000000	0x00002000
<input type="checkbox"/> IO A :		0x00000000	0x00000100
<input type="checkbox"/> IO B :		0x00000000	0x00000100

Buttons: Clear, Write, Cancel

Summit Z3-16 SN:63055

Role	Speed	Link State
Device	2.5 GT/s	???

Summit T3-16 SN:62102

DS	x16	2.5	Speed	Link State
DS	x16	2.5	●●●●●●●●●●	●●●●●●●●●●
US	x16	2.5	●●●●●●●●●●	●●●●●●●●●●

Ready



# SSD Drive Emulation

NVM	R→	2.5	RequesterID	ASQS	ACQS	Time Delta	Time Stamp
4	x4	x4	000:00:0	127	127	1.464 µs	0199 . 362 722 052 s

NVM	R→	2.5	RequesterID	ASQB AddressHi	ASQB AddressLow	Time Delta	Time Stamp
5	x4	x4	000:00:0	0x00000002	0x3D61B000	2.872 µs	0199 . 362 723 516 s

NVM	R→	2.5	RequesterID	ACQB AddressHi	ACQB AddressLow	Time Delta	Time Stamp
6	x4	x4	000:00:0	0x00000002	0x3D61D000	2.904 µs	0199 . 362 726 388 s

NVM	R→	2.5	RequesterID	EN	CSS	MPS	AMS	SHN	IOSQES	IOCQES	Time Delta	Time Stamp
7	x4	x4	000:00:0	1	NVM command set	0	b00	b00	0	0	34.200 µs	0199 . 362 729 292 s

NVM	R→	2.5	RequesterID	CompleterID	RDY	CFS	SHST	Time Delta	Time Stamp
8	x4	x4	000:00:0	008:00:0	1	0	b00	14.988 ms	0199 . 362 763 492 s

NVM	R→	2.5	RequesterID	Admin SQT QID = 0	Time Delta	Time Stamp
9	x4	x4	000:00:0	0x0001	192.276 µs	0199 . 377 751 104 s

NVM	R→	2.5	RequesterID	CompleterID	Admin Cmd	OPC	FUSE	CID	NSID	MPTR Hi	MPTR Low	PRP1 Hi	PRP1 Low	PRP2 Hi	PRP2 Low	CNS
10	x4	x4	008:00:0	000:00:0	Identify	b00	0x0000	0x00000000	0x00000000	0x00000000	0x00000000	0x00000002	0x3D2DA160	0x00000002	0x3D2DB000	Controller

Time Delta	Time Stamp
1.179 ms	0199 . 377 943 380 s

NVM	R→	2.5	RequesterID	Identify CDS	VID	SSVID	SN	MN	FR	RAB	IEEE	MIC	MDTS	OACS	ACL	AERL	FRMW	LPA	ELPE	NPSS	AVSCC	SQES
11	x4	x4	008:00:0	0x1570	0x1570		NVMeLeCroy000000	0x00000000AD55A46B	0	0x0	0	0	0	0x0303	0	6	0	63	0	4	0	0

CQES	NN	ONCS	FUSES	FNA	VWC	AWUN	AWUPF	NVSCC	PSD0	MP	ENLAT	EXLAT	RRT	RRL	RWT	RWL	PSD1	MP	ENLAT	EXLAT	RRT	RRL
0	0x44660000	0x1	0x0	0	2	0x0	0x0	0	0x0000	0x00000000	0x00000000	0x00	0x00	0x00	0x00	0x00	0x0000	0x00000000	0x00000000	0x00	0x00	

RWT	RWL	PSD2	MP	ENLAT	EXLAT	RRT	RRL	RWT	RWL	PSD3	MP	ENLAT	EXLAT	RRT	RRL	RWT	RWL	PSD4	MP	ENLAT	EXLAT	RRT	RRL
0x00	0x00	0x0000	0x00000000	0x00000000	0x00	0x00	0x00	0x00	0x00	0x0000	0x00000000	0x00000000	0x00	0x00	0x00	0x00	0x00	0x0000	0x00000000	0x00000000	0x00	0x00	

RWT	RWL	VS	Data	Time Delta	Time Stamp
0x00	0x00		1023 bytes	663.720 µs	0199 . 379 122 220 s

NVM	R→	2.5	RequesterID	Command Completion	SQHD	SQID	CID	P	ST	SC	SCT	M	DNR	Time Delta	Time Stamp
12	x4	x4	008:00:0	0x00000000	0x0001	0x0000	0x0000	1		0x0000	0x00	0	0	18.232 µs	0199 . 379 785 940 s

Link Tra	R→	2.5	TLP	Mem	MWr(32)	Length	RequesterID	Tag	Address	1st BE	Last BE	Data	VC ID	ExplicitACK	Metrics	# Packets	Time Delta
1965	x4	x4	995		010:00000	1	008:00:0	0	FEE3F00C	1111	0000	1 dword	0	Packet #7599		2	10.420 µs

Time Stamp
0199 . 379 804 172 s

NVM	R→	2.5	RequesterID	Admin CQH QID = 0	Time Delta	Time Stamp
13	x4	x4	000:00:0	0x0001	13.547 ms	0199 . 379 814 592 s



# Setting up Controller Registers

Submission queue size

Submission queue BAR

Completion queue BAR

Enable doorbell execution

Submission queue tail doorbell

NVM	R→	2.5	RequesterID	ASQS	ACQS	Time Stamp							
4	x4		000:00:0	127	127	0199.362722052 s							
Link Tra	R→	2.5	TLP	Mem	MWr(32)	Length	RequesterID	Tag	Address	1st BE	Last BE	Data	
1920	x4		961		010:00000	1	000:00:0	0	FE400024	1111	0000	1 dword	
NVM	R→	2.5	RequesterID	ASQB AddressHi	ASQB AddressLow	Time Stamp							
5	x4		000:00:0	0x00000002	0x3D61B000	0199.362723516 s							
Link Tra	R→	2.5	TLP	Mem	MWr(32)	Length	RequesterID	Tag	Address	1st BE	Last BE	Data	
1921	x4		962		010:00000	1	000:00:0	0	FE400028	1111	0000	00B0613D	
Link Tra	R→	2.5	TLP	Mem	MWr(32)	Length	RequesterID	Tag	Address	1st BE	Last BE	Data	
1922	x4		963		010:00000	1	000:00:0	0	FE40002C	1111	0000	02000000	
NVM	R→	2.5	RequesterID	ACQB AddressHi	ACQB AddressLow	Time Stamp							
6	x4		000:00:0	0x00000002	0x3D61D000	0199.362726388 s							
Link Tra	R→	2.5	TLP	Mem	MWr(32)	Length	RequesterID	Tag	Address	1st BE	Last BE	Data	
1923	x4		964		010:00000	1	000:00:0	0	FE400030	1111	0000	1 dword	
Link Tra	R→	2.5	TLP	Mem	MWr(32)	Length	RequesterID	Tag	Address	1st BE	Last BE	Data	
1924	x4		965		010:00000	1	000:00:0	0	FE400034	1111	0000	1 dword	
NVM	R→	2.5	RequesterID	EN	CSS	MPS	AMS	SHN	IOSQES	IOCQES	Time Stamp		
7	x4		000:00:0	1	NVM command set	0	b00	b00	0	0	0199.362729292 s		
Link Tra	R→	2.5	TLP	Mem	MWr(32)	Length	RequesterID	Tag	Address	1st BE	Last BE	Data	
1925	x4		966		010:00000	1	000:00:0	0	FE400014	1111	0000	1 dword	
NVM	R→	2.5	RequesterID	CompleterID	RDY	CFS	SHST	Time Stamp					
8	x4		000:00:0	008:00:0	1	0	b00	0199.362763492 s					
Split Tra	R→	2.5	Mem	MRd(32)	RequesterID	CompleterID	Tag	TC	VC ID	Address	Status	Data	
959	x4			000:00000	000:00:0	008:00:0	1	0	0	FE40001C	SC	1 dwe	
NVM	R→	2.5	RequesterID	Admin SQT QID = 0	Time Stamp								
9	x4		000:00:0	0x0001	0199.377751104 s								
Link Tra	R→	2.5	TLP	Mem	MWr(32)	Length	RequesterID	Tag	Address	1st BE	Last BE	Data	
1928	x4		968		010:00000	1	000:00:0	0	FE401000	1111	0000	1 dword	



# Identify Command Execution

NVM	2.5	RequesterID	CompleterID	Admin Cmd	OPC	FUSE	CID	NSID	MPTR Hi	MPTR Low	PRP1 Hi	PRP1 Low	PRP2 Hi	PRP2 Low	CNS	Time Stamp
10	R-	x4	008:00:0	000:00:0	Identify	b00	0x0000	0x00000000	0x00000000	0x00000000	0x00000002	0x3D2DA160	0x00000002	0x3D2DB000	Controller	0199.377943380 s

Split Tra	2.5	Mem	MRd(64)	RequesterID	CompleterID	Tag	TC	VC ID	Address	Status
960	R-	x4	001:00000	008:00:0	000:00:0	0	0	0	00000002:3D61B000	SC

Data	Metrics	# LinkTras	Time Delta	Time Stamp
0: 06000000 00000000 00000000 00000000 00000000 00000000 60A12D3D 02000000		2	1.179 ms	0199.377943380 s
8: 00B02D3D 02000000 01000000 00000000 00000000 00000000 00000000 00000000				

NVM	2.5	RequesterID	Identify CDS	VID	SSVID	SN	MN	FR	RAB	IEEE	MIC	MDTS	OACS	ACL	AERL	FRMW	LPA	ELPE	NPSS	AVSCC	SQES	CQES	NN	ONCS
11	R-	x4	008:00:0	0x1570	0x1570		NVMeLeCroy000000	0x000000008D0DB8E1	0	0x0	0	0	0x0303	0	6	0	63	0	4	0	0	0	0x44660000	0x1

FUSES	FNA	VWC	AWUN	AWUPF	NVSCC	PSD0	MP	ENLAT	EXLAT	RRT	RRL	RWT	RWL	PSD1	MP	ENLAT	EXLAT	RRT	RRL	RWT	RWL	PSD2	MP	ENLAT	EXLAT	RRT
0x0	0	2	0x0	0x0	0	0x0000	0x00000000	0x00000000	0x00000000	0x00	0x00	0x00	0x00	0x0000	0x00000000	0x00000000	0x00000000	0x00	0x00	0x00	0x00	0x0000	0x00000000	0x00000000	0x00000000	0x00

RRL	RWT	RWL	PSD3	MP	ENLAT	EXLAT	RRT	RRL	RWT	RWL	PSD4	MP	ENLAT	EXLAT	RRT	RRL	RWT	RWL	VS	Data	Time Stamp
0x00	0x00	0x00	0x0000	0x00000000	0x00000000	0x00	0x00	0x00	0x00	0x00	0x0000	0x00000000	0x00000000	0x00	0x00	0x00	0x00	0x00	0x00	1023 bytes	0199.379122220 s

Link Tra	2.5	TLP	Mem	MW(64)	Length	RequesterID	Tag	Address	1st BE	Last BE	Data	VC ID	Explicit ACK	Metrics	# Packets	Time Delta	Time Stamp
1931	R-	x4	961	011:00000	8	008:00:0	0	00000002:3D2DA160	1111	1111	8 dwords	0	Packet#7531		2	36.208 μs	0199.379122220 s
1932	R-	x4	962	011:00000	32	008:00:0	0	00000002:3D2DA180	1111	1111	32 dwords	0	Packet#7533		2	19.888 μs	0199.379158428 s
1933	R-	x4	963	011:00000	32	008:00:0	0	00000002:3D2DA200	1111	1111	32 dwords	0	Packet#7535		2	23.128 μs	0199.379178316 s
1934	R-	x4	964	011:00000	32	008:00:0	0	00000002:3D2DA280	1111	1111	32 dwords	0	Packet#7537		2	18.512 μs	0199.379201444 s
1935	R-	x4	965	011:00000	32	008:00:0	0	00000002:3D2DA300	1111	1111	32 dwords	0	Packet#7539		2	18.664 μs	0199.379219956 s
1936	R-	x4	966	011:00000	32	008:00:0	0	00000002:3D2DA380	1111	1111	32 dwords	0	Packet#7541		2	18.904 μs	0199.379238620 s

# Identify Command Execution

## System Memory command

Address	Data
0x3D61B000	0x3D2DA160

## NVMe Controller registers

Register	Address	Data
ASQB	0x28	3D61B000

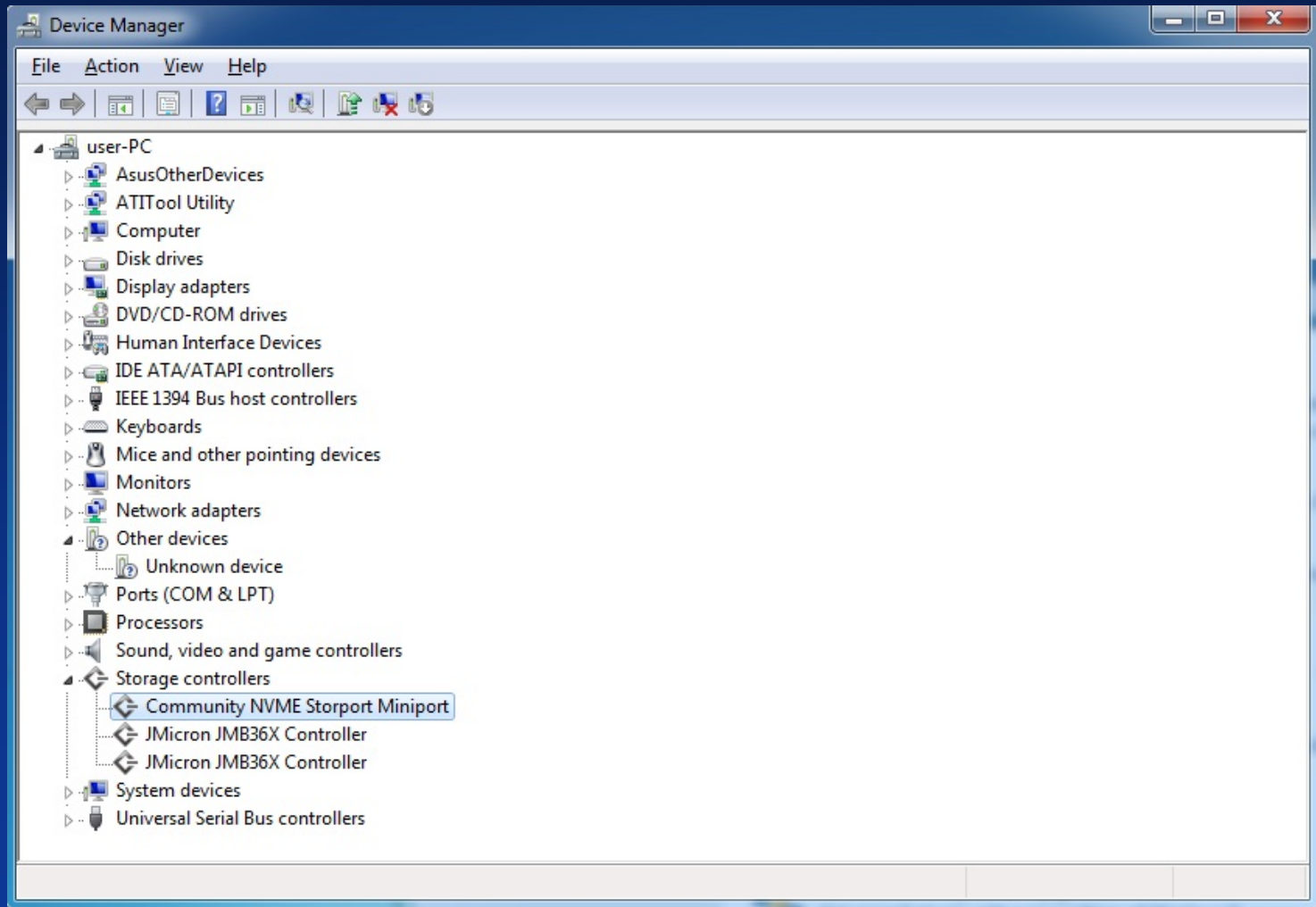
## System Memory Data

Address	Data
0x3D2DA160	0x70157015

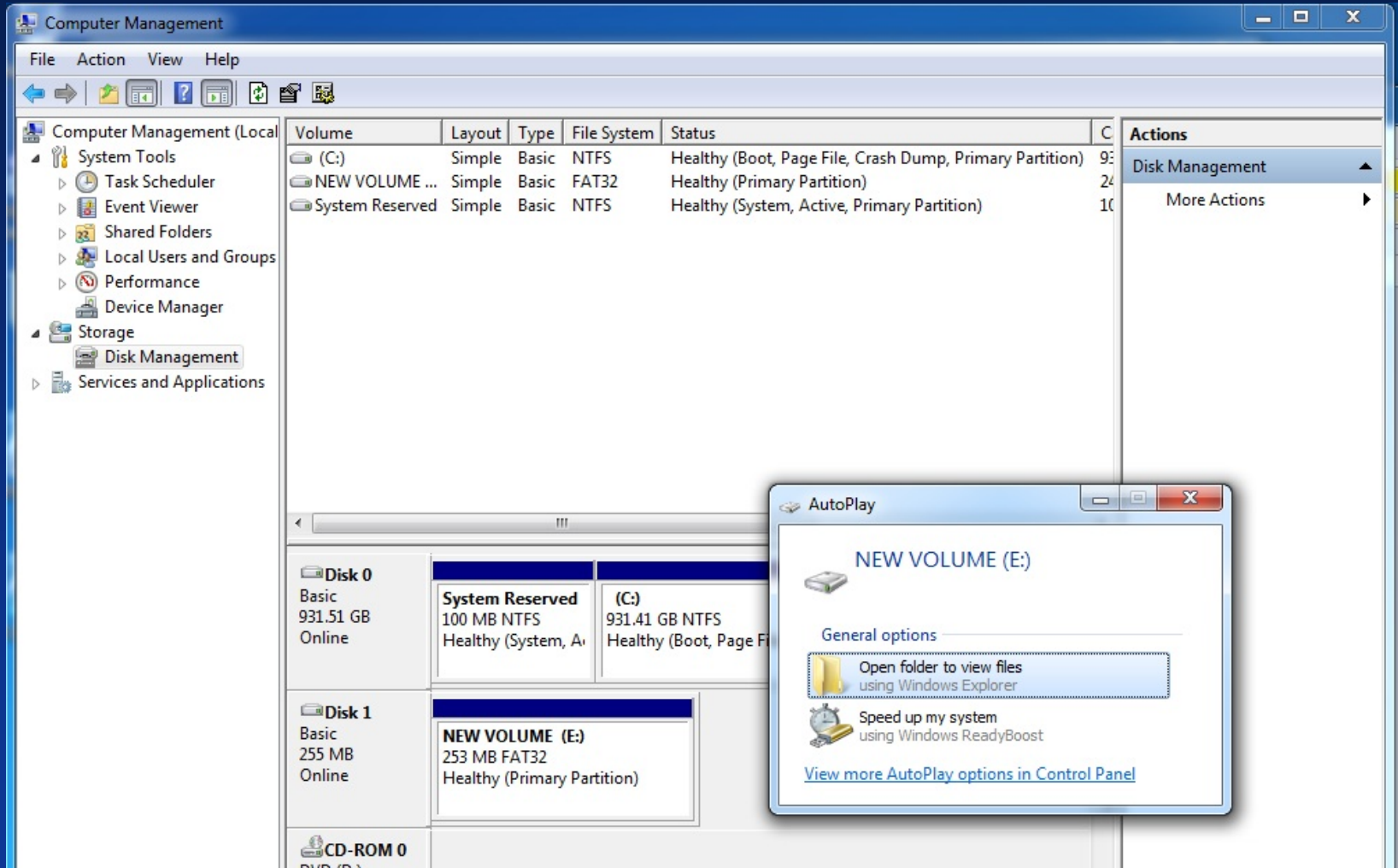
## NVMe Device memory space

Address	Data
0x3D61B000	3D2DA160

# Emulated NVMe Device Shown in Device Manager



# Emulated Drive Shown in Disk Management



The screenshot shows the Windows Computer Management console. The 'Disk Management' section is expanded, showing three disks:

- Disk 0:** Basic, 931.51 GB, Online. Contains 'System Reserved' (100 MB NTFS, Healthy) and '(C:)' (931.41 GB NTFS, Healthy).
- Disk 1:** Basic, 255 MB, Online. Contains 'NEW VOLUME (E:)' (253 MB FAT32, Healthy).
- CD-ROM 0:** DVD (D:)

An AutoPlay dialog box is open for the 'NEW VOLUME (E:)', showing options to 'Open folder to view files using Windows Explorer' and 'Speed up my system using Windows ReadyBoost'. The 'General options' section is highlighted.

Volume	Layout	Type	File System	Status
(C:)	Simple	Basic	NTFS	Healthy (Boot, Page File, Crash Dump, Primary Partition)
NEW VOLUME ...	Simple	Basic	FAT32	Healthy (Primary Partition)
System Reserved	Simple	Basic	NTFS	Healthy (System, Active, Primary Partition)

# Agenda

- High speed protocol convergence
- Storage over PCI Express architectural description
- Command queue generation example
- Emulating an SSD controller
- Emulating an SSD host
- Command Validation

# Testing NVM Express

**nvm**  
EXPRESS



Summit T3-16  
Analyzer

Summit T3-8  
Analyzer

Summit T28  
Analyzer

Summit Z3-16  
Exerciser with  
Test Platform

- NVM Commands
  - ✓ Write
  - ✓ Read
  - ✓ Compare
  - ✓ Extensible for Vendor Specific Commands
- Queue Management
- Come up in Device Manager
- Extensible Vendor Specific Features (for Get/Set Features)
- Complete commands via fused Commands (i.e. Compare & Write)



# Initialization example

**; This script performs basic initialization of an NVMe device**

**; Set up BARs;**

**; Enable bus master, memory space,  
set interrupt disable;**

**; Set Max payload and read request in Device Control;**

**; Write MSI-X table and enable MSI-X.**

```
include="nvme_definitions.peg"
```

```
packet="Temp_ConfigWrite0"
```

```
{
```

```
    Register = 0x10
```

```
    Payload = ( BAR0_ADDRESS_FLIPPED )
```

```
}
```

```
wait=TLP { TLPType = Cpl }
```

```
packet="Temp_ConfigWrite0"
```

```
{
```

```
    Register = 0x14
```

```
    Payload = ( 0 )
```

```
}
```

```
wait=TLP { TLPType = Cpl }
```

**; This script performs NVM specific initialization by writing Controller registers on the device**

```
include="nvme_definitions.peg"
```

```
; Set ACQS and ASQS in AQA – Admin Queue Attributes register
```

```
packet="Temp_OneDwordWrite"
```

```
{
```

```
    Address = ( CONTROLLER_REGISTERS_BASE + 0x24 )
```

```
        Payload = ( 7F007F00 )
```

```
}
```

```
; Set Admin submission Queue address base ASQB high and low . This address corresponds to the base address set for Mem_64 Host region in the generation options file "host_go.gen"
```

```
; ASQ – Admin Submission Queue Base Address low
```

```
packet="Temp_OneDwordWrite"
```

```
{
```

```
    Address =
```

```
    ( CONTROLLER_REGISTERS_BASE + 0x28 )
```

```
    Payload = ( 0080AA2F ) }
```



# Agenda

- Layered Protocol Stack convergence
- Storage over PCI Express architectural description
- Command queue generation example
- Emulating an SSD controller
- Emulating an SSD host
- Command Validation





# Exerciser Features for Storage over PCIe validation

- Read completion payload storage for later processing to implement command queuing
- Branch upon write payload and procedure activation to implement doorbell registers
- DMA descriptor implementation and the use of descriptor data through field substitution
- Creation of data structures in emulator memory
- Trace export to generation file with different timing options
- Extraction of configuration file from trace and import to device emulator



# Command Validation- NVMe - Generation script

```
include="nvme_start.peg"
```

```
; Pre-program command in the Host Memory Region. This is the Mem_64 Host region
```

```
; defined in the generation options file "nvme_host_gen_options.gen"
```

```
; Command data is copied from the trace taken a the last call
```

```
AddressSpace=Write
```

```
{
```

```
    Location=Mem64
```

```
    Offset = 0
```

```
    Size = 128
```

```
    LoadFrom =
```

```
(0x06 0x00 0x00 0x00 0x00 0x00 0x00 0x00 0x00 0x00 0x00 0x00 0x00 0x00 0x00 0x00  
0x00 0x00 0x00 0x00 0x00 0x00 0x00 0x00 0x90 0xB1 0x25 0x3F 0x08 0x00 0x00 0x00  
0x00 0xC0 0x25 0x3F 0x08 0x00 0x00 0x00 0x01 0x00 0x00 0x00 0x00 0x00 0x00 0x00  
0x00 0x00 0x00 0x00 0x00 0x00 0x00 0x00 0x00 0x00 0x00 0x00 0x00 0x00 0x00 0x00)
```

```
0x06 0x00 0x01 0x00 0x01 0x00 0x00 0x00 0x00 0x00 0x00 0x00 0x00 0x00 0x00 0x00  
0x00 0x00 0x00 0x00 0x00 0x00 0x00 0x00 0x00 0xE0 0x47 0x3E 0x02 0x00 0x00 0x00  
0x00 0x00 0x00 0x00 0x00 0x00 0x00 0x00 0x00 0x00 0x00 0x00 0x00 0x00 0x00 0x00  
0x00 0x00 0x00 0x00 0x00 0x00 0x00 0x00 0x00 0x00 0x00 0x00 0x00 0x00 0x00 0x00)
```

```
} # Identify Controller and Identify Namespace
```



# Command Validation- NVMe - Generation script

**; Write Admin Submission queueTail Doorbell Register**

```
packet="Temp_OneDwordWrite"
```

```
{
```

```
    Address = ( CONTROLLER_REGISTERS_BASE + 0x1000 )
```

```
    Payload = ( 01000000 )
```

```
}
```

**; Wait for the Controller to process the command. This will include writing the Identify data,  
; writing the Admin completion Queue, and sending the MSI-X interrupt at vector**

```
wait=TLP
```

```
{
```

```
    TLPType = MWr32
```

```
    Address = ADMIN_INT_VECTOR_ADDRESS
```

```
}
```

**; Write Admin Completion Queue Head Doorbell Register**

```
packet="Temp_OneDwordWrite"
```

```
{
```

```
    Address = ( CONTROLLER_REGISTERS_BASE + 0x1004 )
```

```
    Payload = ( 01000000 )
```

```
}
```



# Command Validation- NVMe - Verification script

## # Constant definitions

```
# Test stage definitions, should be sequential
const STAGE_NVME_CONFIG      = 0;
const STAGE_ADMIN_DORBELL_1  = 1;
const STAGE_READ_CMD_1      = 2;
const STAGE_TRANSFER_DATA_1  = 3;
const STAGE_WRITE_CPL_1     = 4;
const STAGE_SEND_INTERRUPT_1 = 5;
const STAGE_ADMIN_DORBELL_2  = 6;
const STAGE_READ_CMD_2      = 7;
const STAGE_TRANSFER_DATA_2  = 8;
const STAGE_WRITE_CPL_2     = 9;
const STAGE_SEND_INTERRUPT_2 = 10;
```

## # Variable declarations

```
set Admin_SQB_Low = 0;
set Admin_SQB_High = 0;
set Admin_CQB_Low = 0;
set Admin_CQB_High = 0;
set PRP1_High = 0;
set PRP1_Low = 0;
set PRP2_High = 0;
set PRP2_Low = 0;
set Cmd_Dw10 = 0;
set CurrentIdentifyXferredLength = 0;
set TestStage = 0;
set CurrentChannel = 0;
```



# Command Validation- NVMe - Verification script

# Function: OnStartScript()

# **Description:** The application calls this function at the beginning of the script execution.

**OnStartScript()**

```
{ ReportText( "Verifying Identify Command..." );  
  SendAllChannels();  
  SendLevelOnly( _LINK );  
  SendTraceEvent( _LINK_CONFIG );  
  SendTraceEvent( _LINK_COMPLETION );  
  SendTraceEvent( _LINK_MEMORY );  
  Admin_SQB_Low = 0; # initialize variables  
  Admin_SQB_High = 0;  
  Admin_CQB_Low = 0;  
  Admin_CQB_High = 0;  
  PRP1_High = 0;  
  PRP1_Low = 0;  
  PRP2_High = 0;  
  PRP2_Low = 0;  
  Cmd_Dw10 = 0;  
  CurrentIdentifyXferredLength = 0;  
  TestStage = STAGE_NVME_CONFIG;}
```



# Command Validation- NVMe- Verification script

```
# Function: ProcessEvent()
# Description: Entry point of the script.
# The application calls this function every time it finds the relevant trace event.
ProcessEvent()
{
    CurrentChannel = in.Channel;
    event_type = in.TraceEvent;
    # transaction status checking
    if( in.TransactionStatus == LINK_TRA_STATUS_INCOMPLETE )
    {
        FailTest( "Transaction wasn't complete at the Link Layer" );
        return null;
    }
    select
    {
        event_type == _LINK_CONFIG      : ProcessCfgRequest();
        event_type == _LINK_COMPLETION  : ProcessCompletion();
        event_type == _LINK_MEMORY     : ProcessMemReadOrWrite();
    };
    return Complete();
}
```



# NVMe Command Validation Resulting Trace

Split Tra	R	2.5	Cfg	CfgRd0	RequesterID	CompleterID	Tag	TC	VC ID	DeviceID	Register	Status	Class Code	Revision ID	Metrics	# LinkTras	Time Delta	Time Stamp						
9	x1	x1	000:00100	064:02:0	129:00:0	0	0	0	0	129:00:0	0x008	SC	0x010802	0x01	2	165.760 us	0002 . 746 951 600 s							
Split Tra	R	2.5	Cfg	CfgRd0	RequesterID	CompleterID	Tag	TC	VC ID	DeviceID	Register	Status	Base Address Register 0	Metrics	# LinkTras	Time Delta	Time Stamp							
10	x1	x1	000:00100	064:02:0	129:00:0	0	0	0	0	129:00:0	0x010	SC	0xEC030004	2	159.864 us	0002 . 747 117 360 s								
NVM	R	2.5	RequesterID	CC	EN	CSS	MPS	AMS	SHN	IOSQES	IOCQES	Time Delta	Time Stamp											
0	x1	x1	064:02:0	0	0	NVM command set	0	b00	b00	0	0	26.888 us	0002 . 747 277 224 s											
NVM	R	2.5	RequesterID	AQA	ASQS	ACQS	Time Delta	Time Stamp																
1	x1	x1	064:02:0	127	127	18.936 us	0002 . 747 304 112 s																	
NVM	R	2.5	RequesterID	ASQ	ASQB AddressHi	ASQB AddressLow	Time Delta	Time Stamp																
2	x1	x1	064:02:0	0x00000004	0x2FAA8000	40.600 us	0002 . 747 323 048 s																	
NVM	R	2.5	RequesterID	ACQ	ACQB AddressHi	ACQB AddressLow	Time Delta	Time Stamp																
3	x1	x1	064:02:0	0x00000004	0x2FAAA000	37.608 us	0002 . 747 363 648 s																	
NVM	R	2.5	RequesterID	CC	EN	CSS	MPS	AMS	SHN	IOSQES	IOCQES	Time Delta	Time Stamp											
4	x1	x1	064:02:0	1	0	NVM command set	0	b00	b00	0	0	775.660 ms	0002 . 747 401 256 s											
NVM	R	2.5	RequesterID	SQyTDBL	Admin SQT QID = 0	Time Delta	Time Stamp																	
5	x1	x1	064:02:0	0x0001	704.000 ns	0003 . 523 061 552 s																		
NVM	R	2.5	RequesterID	CompleterID	Admin Cmd	OPC	FUSE	CID	NSID	MPTR Hi	MPTR Low	PRP1 Hi	PRP1 Low	PRP2 Hi	PRP2 Low	CNS	Time Delta	Time Stamp						
6	x1	x1	129:00:0	000:00:0	Identify	b00	0x0000	0x00000000	0x00000000	0x00000000	0x00000000	0x00000008	0x3F25B190	0x00000008	0x3F25C000	Controller	275.656 us	0003 . 523 062 256 s						
NVM	R	2.5	RequesterID	Identify CDS	VID	SSVID	SN	MN	FR	RAB	IEEE	MIC	MDTS	OACS	ACL	AERL	FRMW	LPA	ELPE	NPSS	AVSQC	SQES	CQES	NN
7	x1	x1	129:00:0	0x111D	0x1D11	IDT_SN_0000	IDT V.Board NVMe SSD	1.5.0	1	0x313233	1	0	0x0006	3	5	5	1	63	0	1	102	68	0x00000001	
			ONCS	FUSES	FNA	VWC	AWUN	AWUPF	NVSCC	PSD0	MP	ENLAT	EXLAT	RRT	RRL	RWT	RWL	vs	Data	Time Delta	Time Stamp			
			0x0006	0x0000	0	1	0xFFFF	0xFFFF	1	0x0BB8	0x00000064	0x00000064	0x00	0x00	0x00	0x00	1024 bytes	2.007 ms	0003 . 523 337 912 s					
NVM	R	2.5	RequesterID	Command Completion	SQHD	SQID	CID	P	ST	SC	SCT	M	DNR	Time Delta	Time Stamp									
8	x1	x1	129:00:0	0x00000000	0x0001	0x0000	0x0000	1	0x0000	0x00	0	0	76.352 us	0003 . 525 344 736 s										
Link Tra	R	2.5	TLP	Mem	MWr(32)	Length	RequesterID	Tag	Address	1st BE	Last BE	Data	VC ID	ExplicitACK	Metrics	# Packets	Time Delta	Time Stamp						
66	x1	x1	1070	010:00000	1	129:00:0	0	FFABF00C	1111	0000	1 dword	0	Packet#133	2	925.880 us	0003 . 525 421 088 s								
NVM	R	2.5	RequesterID	CQyHDBL	Admin CQH QID = 0	Time Delta	Time Stamp																	
9	x1	x1	064:02:0	0x0001	161.856 us	0003 . 526 346 968 s																		
NVM	R	2.5	RequesterID	SQyTDBL	Admin SQT QID = 0	Time Delta	Time Stamp																	
10	x1	x1	064:02:0	0x0002	728.000 ns	0003 . 526 508 824 s																		
NVM	R	2.5	RequesterID	CompleterID	Admin Cmd	OPC	FUSE	CID	NSID	MPTR Hi	MPTR Low	PRP1 Hi	PRP1 Low	PRP2 Hi	PRP2 Low	CNS	Time Delta	Time Stamp						
11	x1	x1	129:00:0	000:00:0	Identify	b00	0x0001	0x00000001	0x00000000	0x00000000	0x00000000	0x00000002	0x3E47E000	0x00000000	0x00000000	Namespace	211.320 us	0003 . 526 509 552 s						

- Storage devices have adopted the serial protocol host interface
- SSDs are becoming an integral part of the Enterprise infrastructure
- SSDs are moving towards a PCIe host interface
- NVM Express, SCSI Express, and SATA Express are new compelling implementations for the SSD host interface
- PCI Express based SSD implementations leverage off PCI Express Analysis expertise